AU8936908

(12) PATENT ABRIDGMENT (11) Document No. AU-B-36908/89
(19) AUSTRALIAN PATENT OFFICE (10) Acceptance No. 633830

(54) Title
METHOD FOR HLA DP TYPING

International Patent Classification(s)
(51)[4] C12Q 001/68 C07H 021/04
(51)[5] G01N 033/564

(21) Application No. : 36908/89 (22) Application Date : 18.05.89

(87) PCT Publication Number : WO89/11547

(30) Priority Data

| (31) Number | (32) Date | (33) Country |
|---|---|---|
| 196660 | 20.05.88 | US UNITED STATES OF AMERICA |
| 258212 | 14.10.88 | US UNITED STATES OF AMERICA |
| 347506 | 04.05.89 | US UNITED STATES OF AMERICA |

(43) Publication Date : 12.12.89

(44) Publication Date of Accepted Application : 11.02.93

(71) Applicant(s)
CETUS CORPORATION

(72) Inventor(s)
HENRY A. ERLICH; GLENN T. HORN; TEODORICA L. BUGAWAN

Attorney or Agent
SPRUSON & FERGUSON , GPO Box 3898, SYDNEY NSW 2001

(56) Prior Art Documents
EP 237362

(57) Claim

1. A process for determining an individual's HLA DPbeta genotype from a nucleic acid containing sample obtained from the individual comprising:

a) amplifying a sequence from exon two of a DPbeta gene in the nucleic acids in the sample;

b) mixing the amplified nucleic acids with a panel of sequence specific oligonucleotide (SSO) probes, wherein each probe comprises a nucleotide sequence complementary to a variant sequence of a variable segment of said HLA DPbeta gene wherein the variant region is selected from the group consisting of the variable segments located at codons 8-11, 36, 55-57, 65-69, 76 and 84-87, under conditions wherein SSO probes bind to said amplified nucleic acids to form stable hybrid duplexes only if they are exactly complementary; and

c) detecting hybrids formed between the amplified nucleic acids and the SSO probes.

8. Sequence specific oligonucleotide (SSO) probes useful for determining an individual's HLA DPbeta genotype from a nucleic acid containing sample obtained form the individual, each of said probes comprising a nucleotide sequence that is totally complementary to a variant sequence of a variable segment selected from the group consisting of the variable segments located at codons 8-11, 36, 55-57, 65-69, 76 and 84-87 of the second exon of a DPbeta allele.

AUSTRALIA

Patents Act

COMPLETE SPECIFICATION
(ORIGINAL)

Class Int. Class

Application Number:
Lodged:

Complete Specification Lodged:
Accepted:
Published:

Priority

Related Art:

---

Applicant(s):

Imperial Chemical Industries PLC
Imperial Chemical House, Millbank, London SW1P 3JF, UNITED KINGDOM

Address for Service is:

PHILLIPS ORMONDE & FITZPATRICK
Patent and Trade Mark Attorneys
367 Collins Street
Melbourne 3000 AUSTRALIA

Complete Specification for the invention entitled:

**EXTENDED NUCLEOTIDE SEQUENCES**

Our Ref : 154304
POF Code: 1453/1453

The following statement is a full description of this invention, including the best method of performing it known to applicant(s):



6006

TITLE: EXTENDED NUCLEOTIDE SEQUENCES

The present invention relates generally to a method of characterising a genomic DNA sample and to nucleotide sequences employed in the method. In particular the invention involves the use of nucleotide sequences comprising oligonucleotides hybridisable to regions adjacent informative genetic loci. The method of the invention may for example be used in paternity disputes, forensic medicine or in the prevention, diagnosis and treatment of genetic disorders or predispositions. The method is of particular use where only as little as one molecule of an informative locus is present in the genomic DNA sample.

Methods of genetic characterisation are known in the art. In UK patent no. 2166445 (Lister Institute) there are described various DNA sequences which may be used as probes to hybridise simultaneously to a number of polymorphic sites within animal, e.g. human, and plant genomes enabling the production of a "DNA fingerprint" composed of marked DNA bands of differing molecular weights. The DNA fingerprint as a whole is characteristic of the individual concerned and the origin of the differing bands can be traced through the ancestry of the individual and can in certain cases be postulated as associated with certain genetic disorders. In European Patent Application, Publication No. 238329, there are described various DNA sequences which may be used as probes to hybridise individually at individual polymorphic sites within the animal, for example human genome. Methods of genetic characterisation using one or more of such probes are described.

Tandem-repetitive minisatellite regions in vertebrate DNA frequently show high levels of allelic variability in the number of repeat units [1-4]. Hybridization probes capable of detecting multiple minisatellites and producing individual-specific DNA fingerprints have been developed [5-7], as well as cloned human minisatellites which provide locus-specific probes for individual hypervariable loci [5, 8-10]. These highly informative genetic markers have found widespread application in many areas of genetics, including linkage analysis [9,

11-13], determination of kinship in for example paternity and immigration disputes [6, 10, 14, 15], monitoring bone marrow transplants [16,17], and for individual identification in forensic medicine [10, 18, 19]. Applications to typing forensic samples such as blood and semen strains or hair roots are however limited by the sensitivity of the hybridization probes, which require at least 50ng of relatively undegraded human DNA for typing with locus-specific minisatellite probes [10] and 0.01- g DNA for analysis with multilocus DNA fingerprint probes [6].

Where a sufficient amount of sample DNA is available in the test sample the above disclosures provide valuable and reliable methods of genetic characterisation. However the efficiency of the above techniques is reduced where the amount of genomic DNA in the test sample is limited, for example in forensic applications where often only small copy numbers of the test DNA molecule are available.

It is therefore desirable to provide a further method of genetic characterisation which is especially suitable for small samples of genomic DNA.

K. Kleppe et al in J. Mol.Biol. (1971), 56, 341-361 disclose a method for the amplification of a desired DNA sequence. The method involves denaturation of a DNA duplex to form single strands. The denaturation step is carried out in the presence of a sufficiently large excess of two nucleic acid primers which hybridise to regions adjacent the desired DNA sequence. Upon cooling two structures are obtained each containing the full length of the template strand appropriately complexed with primer. DNA polymerase and a sufficient amount of each required nucleoside triphosphate are added whereby two molecules of the original duplex are obtained. The above cycle of denaturation, primer addition and extension are repeated until the appropriate number of copies of the desired DNA sequence is obtained. It is indicated that adjustment of the primer concentration may be required.

The above method is now referred to as polymerase chain reaction (PCR) as claimed in United States patents nos. 4683195 and 4683202 wherein amplification of a given nucleic acid sequence on a template is effected by extension of a nucleic acid primer in the presence of Taq. polymerase or the Klenow fragment of E.coli DNA polymerase I. The amplification procedure is generally repeated for up to about 50 cycles. The examples provided only relate to short DNA sequences, generally of a few hundred base pairs.

The enzymatic amplification of DNA by the polymerase chain reaction (PCR, [20]) enables such smaller amounts of human DNA to be analysed. The remarkable specificity of thermostable Taq polymerase has greatly simplified PCR [21] and has allowed typing of some classes of human DNA polymorphism to be extended to single hair roots [22] and indeed to individual somatic cells and sperm [23]. In most work to date, PCR has been used to amplify short regions of human DNA, usually a few hundred base pairs long [21-23]. Base substitutional polymorphisms can be detected by hybridizing PCR products with allele-specific oligonucleotides [22, 23], by DNA sequence analysis of PCR products [24], or, if the base substitution affects a restriction site, by cleavage of PCR products with a restriction endonuclease [25]. Deletion/insertion polymorphisms can likewise be analysed by sizing PCR products by gel electrophoresis [22]. Most of these marker systems are however dimorphic and their utility in for example forensic medicine is limited by their relatively low variability in human populations.

As explained above PCR has been restricted to the amplification of relatively short DNA sequences and there are serious doubts as to whether PCR may be employed successfully in relation to long DNA sequences especially if faithful reproduction of such sequences is necessary. Indeed 2kb has been stated to represent the absolute limit of PCR [21]. Further difficulties arise however where the sample DNA contains repetitive sequences, for example tandem repeats of a particular core or consensus sequence such as those found at highly informative hypervariable loci.

| Cell line | ATCC Deposit No. |
|---|---|
| EB-3 | CCL 85 |
| RAJI | CCL 83 |
| JIYOYE (P-2003) | CCL 87 |
| WI-26 | CCL 95 |
| Detroit 551 | CCL 110 |
| RPMI 6666 | CCL 113 |
| RPMI 7666 | CCL 114 |
| CCRF-CEM | CCL 119 |
| CCRF-SB | CCL 120 |
| HT-1080 | CCL 121 |
| HG 261 | CCL 122 |
| CHP3 (M.W.) | CCL 132 |
| LL47 (MaDo) | CCL 135 |
| HEL 299 | CCL 137 |
| LL 24 | CCL 151 |
| HFLI | CCL 153 |
| WI-1003 | CCL 154 |
| MRC-5 | CCL 171 |
| IMR-90 | CCL 186 |
| LS 174T | CCL 188 |
| LL 86(LeSa) | CCL 190 |
| LL 97A (AIMy) | CCL 191 |
| HLF-a | CCL 199 |
| CCD-13Lu | CCL 200 |
| CCD-8Lu | CCL 201 |
| CCD-11Lu | CCL 202 |
| CCD-14Br | CCL 203 |
| CCD-16Lu | CCL 204 |
| CCD-18Lu | CCL 205 |
| CCD-19Lu | CCL 210 |
| Hs888Lu | CCL 21 |
| MRC-9 | CCL 212 |
| Daudi | CCL 213 |
| CCD-25Lu | CCL 215 |

| Cell line | ATCC Deposit No. |
|---|---|
| SW403 | CCL 230 |
| NAMALWA | CRL 1432 |

All the above-mentioned cell lines are freely available from the ATCC. 12301 Parklawn Drive, Rockville, Maryland 20852-1776, USA and are listed in the ATCC catalogue of Cell Lines and Hybridomas. All the above-mentioned cell lines were on deposit with the ATCC prior to 1985.

Convenient informative loci for use in the present invention include loci to which nucleotide sequences and probes disclosed in European patent application no. 238329 are capable of hybridisation, as well as those informative loci the flanking sequences of which are disclosed in this application. Further informative loci include genomic DNA regions identified by minisatellite probes disclosed by S J Gendler et al. in PNAS, 84, (1987), pages 6060-6064. A particularly informative locus is the 5' alpha globin HVR disclosed by the American Journal of Human Genetics, 1988, 43, pages 249-256.

Convenient informative loci for use in the present invention include those of up to 15 kilobases. More convenient informative loci include those which provide short alleles, the upper limit being for example of less than 10 kilobases particularly of less than 8, more particularly less than 6 kilobases. Conveniently the alleles are of at least 1, particularly at least 1.5, more particularly at least 2 kilobases. Thus suitable ranges include 1, 1.5, 2, 2.5, or 3, to 6, 7, 8, 9 or 10 kilobases. In respect of the performance of the method of the invention starting with only a single sample DNA molecule it is preferred that the length of the alleles does not exceed 6 kilobases. Also preferred are informative loci with a restricted range of allele length. It has been observed that these preferences are not always compatible with highly informative loci since these are usually associated with large numbers of minisatellite repeat units and long

alleles. The skilled man will however have no difficulty in selecting convenient loci for his purpose.

Primers capable of hybridisation to regions of sample DNA flanking informative loci are required to provide points for the initiation of synthesis of extension products across informative genetic loci. Such primers are generally oligonucleotides and are preferably single stranded for maximum efficiency in amplification, but may alternatively be double stranded. If double stranded, the primer is first treated to separate its strands before being used to prepare extension polynucleotides. Primers must be sufficiently long to prime the synthesis of extension polynucleotides. The exact length of such primers will depend on many factors including temperature and the nature of the primer. Generally a primer will comprise at least seven nucleotides, such as 15-25 nucleotides, for example 20-25 nucleotides. It will be appreciated that the flanking sequence need not reflect the exact sequence of the flanking region of the informative locus. It is merely necessary that the respective sequences are homologous to a degree enabling hybridisation. The extension product so obtained must also be capable of acting as a template for further hybridisation with a primer. The maximum length of any primer is not believed to be critical and is only limited by practical considerations.

Normally the test sample DNA will be double stranded and it is therefore convenient to select primers which hybridise to different strands of the sample DNA and at relative positions along the sequence such that an extension product synthesised from one primer can act as a template for extension of the other primer into a nucleic acid of defined length.

The flanking regions of certain of the informative genetic regions identified by the probes disclosed in UK Patent No. 2166445 and European patent application no. 238329 referred to above and nucleotide sequences hybridisable thereto have not been previously disclosed and both individually and in any combination represent further aspects of the present invention. The sequences are set out

on pages preceding the Examples. The informative genetic regions are indicated on some sequences by the word MINISATELLITE. As explained above any convenient sequence may be used as a primer. Examples of some convenient primer sequences have been underlined. For clarity and brevity, on some sequences only the outermost repeat unit on each side of each minisatellite tandem array is shown (in capitals), separated by a series of x's to indicate the omission of most repeats. Since there is seldom a whole number of repeat units in each array, these outermost repeats may be out of register with each other, but preserve the correct relation to the immediate flanking DNA.

A previously unsuspected tandem repeat region was found by sequencing pMS31 (EP-238329), this array is separated from the major minisatellite by 10 bases, and consists of more than 7 repeats of a 19bp G/C rich sequence which extends to the Sau3AI site defining the end of the clone. This region, designated 31B, is G-rich on the opposite strand to the G-rich strand of the major minisatellite. Genomic mapping shows that, if variable at all, 31B makes a minimal contribution to the length variation at this locus (data not shown).

Two loci were detected by pMS228 (Armour et al. 1989, NAR, 17, 13, 4925-4935) at high stringency. Family analyses showed these loci to be tightly linked. Restriction mapping of the cloned DNA, subsequently confirmed by sequence analysis, demonstrated the presence of two distinct tandem repetitive regions, designated 228A and 228B (Figure 8). Reprobing human DNA with subclones from pMS228 showed that 228A detected the larger, more intensely hybridising locus, and 228B the smaller, more faintly hybridising alleles. These ministallites were localised, using somatic cell hybrids and in situ hybridisation, to 17p13-pter. The properties of the loci detected by pMS51 and pMS228, and by other minisatellites are summarised in Table 2.

Previous evidence for the disposition of a pair of minisatellites in pMS43 as described in our UK patent application 8813781.5 was confirmed by sequence data, and two further examples (pMS31 and

pMS228) of minisatellite clones containing two closely adjacent tandem repeat arrays were elucidated. Unlike pMS43, in which the 'minor' locus 43B has a low (30%) heterozygosity, both the minisatellites in pMS228 have heterozygosities greater than 80%. The minisatellite 228B (figure 8, table 2) combines a high heterozygosity (85%) with a limited allele size range (0.6 -5.5kb); this combination makes it a very useful locus for analysis of minisatellites by the polymerase chain reaction [23]. In general, the most variable loci have a wide range of allele sizes such that many alleles will exceed the maximum size currently amplifiable by this method, so giving an incomplete profile. At the locus detected by 228B, in contrast, a survey of 48 unrelated people showed that 95% of alleles were smaller than 2kb, and that the largest (5.5kb) was well within the amplifiable range (J Armour and A Jeffreys, unpublished). We have shown that even the largest allele at this locus can be amplified, providing the prospect of a complete and yet usefully informative profile from amplification at a single locus.

A novel cloned minisatellite termed MS29 has been isolated which unusually detects two variable loci in human DNA. One locus, located in the terminal region of the short arm of human chromosome 6, is also present in great apes. The second minisatellite locus is located interstitially on chromosome 16p11, and is absent both from non-human primates and from some humans. MS29 was isolated from a L47 genomic library made from DNA enriched for minisatellite sequences as described elsewhere (Wong et al, Ann.Hum.Genet. (1987) 51, 260-288. MS29 comprises a 39bp repeat sequence. This consists of 13 diverged repeats of the trinucleotide YAG wherein Y represents any one of A, G, C or T. The flanking sequences of this minisatellite can be elucidated using sequencing techniques known per se.

In summary the relevant novel flanking sequences are those of MS1, MS29, MS31A and MS31B, MS32, MS43A and MS43B, MS51, MS228A and MS228B. In addition we now also provide further novel flanking sequence information for the minisatellite probe p g3 (Wong et al, Nucleic Acids Research, 14, 11, (1986), 4605-4616). as well as for the probes

33.1, 33.4 and 33.6 (UK Patent 2166445/The Lister Institute of Preventive Medicine).

A particular group of flanking sequences are those of the minisatellite probes MS1, MS29, MS31A and MS31B, MS32, MS43A and MS43B, MS51, MS228A and MS228B.

Further particular groups of flanking sequences are comprised by those of any of the above minisatellite probes.

As stated previously the minisatellites MS29 and MS31B are novel and the repeat sequences and/or flanking sequences of either of these minisatellites represent further particular aspects of the invention.

The method of the present invention can be used in respect of any convenient informative locus. Thus, for example primers may be prepared for hybridisation to the flanking sequences of for example the hypervariable region 5' to the human insulin gene (Am.J.Hum Genet, 1986, 39, 291-229), of the 5' alpha globin HVR (Am.J.Hum. Genet, 1988, 43, 249-256), of the alpha globin 3' HVR (EMBOJ, 1986, 5, 1957-1863), of the hypervariable region at the zeta globin locus (PNAS, 1983, 80, 5022-5026), or of the Ha Ras locus (Nature, 1983, 302, 33-37). Also incorporated by reference in this application are a panel of probes proposed by Ray White. The flanking sequences of these may be elucidated using known techniques.

The extended flanking polynucleotides prepared according to the method of the invention represent a further aspect of the invention. These polynucleotides may be present in single or multiple copies. Where multiple copies are present these are faithful copies and preferably substantially free from networking or cross linkage between individual strands. By the term faithful we mean that the genetic characterisation information to be obtained from the size and composition of the copy is essentially the same for all copies. Conveniently the extended primers comprise a sequence identical with or complementary to an informative genetic locus of more than 1

kilobase. Other convenient values include 1.5 and 2 kilobases. Conveniently the number of copies represent the products of at least 3, 5, 7, 9, 13 or 15 cycles of the amplification method of the present invention.

According to a still further aspect of the invention there is provided a mixture containing multiple faithful copies of extended primers (as hereinbefore defined). As above the mixture is substantially free from networking or cross-linkage between individual strands.

The primers for use in the present invention may be prepared by methods analogous to those known in the art. For example where a given flanking sequence is known, a convenient nucleotide primer may be prepared by direct synthesis. Cloning techniques may also be used to reproduce DNA fragments containing sequences flanking informative genetic loci.

Alternatively DNA fragments comprising informative genetic loci may be identified and using a procedure analogous to that of the methods outlined above, a nucleotide sequence hybridisable to the informative genetic locus may be extended. The product so obtained may then be directly modified, for example by cleavage, to prepare a convenient primer or the sequence thereof may be determined and convenient primers then prepared by direct synthesis.

As explained above the sample DNA is generally double stranded. It is therefore desirable to separate the strands of the nucleic acid before it can act as a template for extension of a primer. Strand separation can be accomplished by any suitable method including physical, chemical or enzymatic means. Conveniently heat denaturation is used to provide up to about 99% denaturation. Typical temperatures used include 85-105°C for times ranging from about 1-10 minutes.

A primer is preferably used for each unique strand of the sample DNA. In respect of double stranded DNA two primers are therefore normally used. Generally the primers are chosen to allow extension to proceed

across the informative locus in the same direction along both strands, that is to say 5' to 3' or vice versa. Preferably extension proceeds 5' to 3'.

Primers are conveniently hybridised with the genomic DNA sample under known conditions. Generally this is allowed to take place in a buffered aqueous solution, preferably at a pH of 7-9. Preferably a large excess of primer is present in the reaction mixture. It will be appreciated that in certain diagnostic applications the amount of sample DNA may not be known but an excess of primers is desirable.

The amplification is conveniently effected by adding the deoxyribonucleoside triphosphates dATP, dCTP, dGTP, and dTTP to the synthesis mixture in adequate amounts and the resulting solution may then, for example be heated to for example about 90°-100°C for a period of for example from about 1 to 10 minutes. After this heating period the solution may be allowed to cool to 30-70°C which is preferable for the primer hybridisation. Where Taq polymerase is used as the inducing agent to facilitate the extension reaction it is preferable to effect primer hybridisation at a temperature of from 50-70°C for example 50°-65°C, such as 60°C.

It has been found that primer hybridisation is advantageously effected in a buffer of reduced ionic strength and at an elevated annealing temperature. This reduces the possibility of mispriming. The expressions "reduced ionic strength" and "elevated annealing temperature" are to be understood as referring to conditions at the extremes of, or outside those ranges which would generally be considered appropriate by the molecular biologist of ordinary skill for a given nucleic acid amplification reaction. By way of illustration, but not limitation a convenient buffer will be of sufficient ionic strength to enable a stable pH of 7-9 to be maintained. Convenient annealing temperatures are generally 50-60°C. It will be appreciated that the actual reaction conditions will depend on the primer(s) used.

An appropriate agent for inducing or catalysing the extension reaction may then be added to the annealed mixture and the reaction may then be allowed to proceed under conditions known in the art.

The inducing agent is conveniently an enzyme. Suitable enzymes include the Klenow fragment of E. coli DNA polymerase I, T4 DNA polymerase, but particularly Taq polymerase. Other available DNA polymerases, reverse transcriptase and other enzymes, including heat stable enzymes which will facilitate the extension reaction may also be used.

The newly synthesised strand and its complementary nucleic acid strand form a double-stranded molecule which is used in the succeeding steps of the method. The steps of strand separation and sequence extension may then be repeated as required using the procedures outlined above. Suitable conditions for such procedures are outlined in US patent nos. 4683202 and 4683195 incorporated by reference above. As explained in the above patent specifications the separation and extension cycles may be performed stepwise or advantageously a plurality of cycles are operated for example in a semi-automated or fully automated manner.

In the conventional PCR technique the amplification reaction is generally allowed to proceed for up to about 50 cycles. In contrast, acccording to the method of the present invention the upper limit for amplification is believed to be about 25 cycles when starting from a single copy of the sample DNA. Where larger starting amounts of genomic DNA sample are available fewer cycles of amplification are required. The convenient window of amplification cycles depending on the amount of sample DNA and the nature of the sequence will now be illustrated, but in no way limited, by reference to the following analysis:

The following analysis refers to amplification reactions using 10$\mu$l volumes and, with extension times of 15 minutes: Considering an individual heterozygous for alleles A and B, where A is longer than B it can be predicted that A will therefore amplify less

efficiently than B. The lower limit of cycle number $C_l$ is dictated by the sensitivity of the probes, which can detect about 0.1 pg product. To detect allele A, sufficient cycles are needed to generate >0.1 pg allele A product. Similarly, >0.1 pg allele B must be produced to detect both alleles. The upper limit of cycle number $C_u$ is limited by the yield of both alleles A and B.

$C_l$ and $C_u$ can be estimated as follows:

Let M = initial mass of human genomic DNA (in picograms).

a = length of allele A (kb)

b = length of allele B (kb)

ga = gain per PCR cycle of allele A

gb = gain per PCR cycle of allele B

Since the diploid genome size of man is $6 \times 10^6$ kb then the initial amount of allele A is

$$M. \frac{a}{6 \times 10^6} \text{ pg}$$

Similarly, the intial amount of allele B is

$$M. \frac{b}{6 \times 10^6} \text{ pg}$$

The yield of allele A after C cycles is given by:

$$M. \frac{a}{6 \times 10^6} g_a^C \text{ pg}$$

and similarly for allele B.

From Fig. 3 the gain per cycle g is related to allele length L by

$$gL = 2 - 0.093\ L \quad (L = 0\text{-}6 \text{ kb})$$

The lower limit of cycles $C_l$ is therefore set by the lowest value of C where

$$\frac{M.a}{6 \times 10^6} (2 - 0.093.L)^C > 0.1$$

and $$\frac{M.b}{6 \times 10^6} (2 - 0.093.L)^C > 0.1$$

The upper limit $C_u$ is likewise set by the highest value of C where the total yield of both alleles is < 1000 pg, i.e.

$$\frac{M.a}{6 \times 10^6} (2 - 0.093.L)^C + \frac{M.b}{6 \times 10^6} (2 - 0.093.L)^C < 1000$$

$C_l$ and $C_u$ can be determined by computer reiteration. Some typical examples are:

| | M | a | b | $C_l$ | $C_u$ |
|---|---|---|---|---|---|
| | pg | kb | kb | | |
| | 10,000 | 5 | 1 | 7 | 20 |
| | 1,000 | 3 | 2 | 10 | 24 |
| | 100 | 6 | 0.5 | 19 | 27 |
| * | 6 | 6 | 1 | $27^+$ | $32^+$ |

+ note that the window becomes very narrow and can disappear for very small amounts of DNA with widely differing allele sizes.

* N.B. This model does not take into account the fact that significant stochastic variation in the number of target molecules exists in amounts of human genomic DNA <100pg (equivalent to 17 cells). The analysis for 6 pg DNA corresponds to the amplification of a single target molecule of each allele.

Homozygotes can be readily accomodated in the model. $C_u$ remains unaltered, whereas $C_l$ is defined by the number of cycles

required to give >0.05 pg of <u>each</u> allele (i.e. >0.1 pg combined).

Reactions with different volumes can also be accomodated. $C_1$ is defined by the number of cycles required to give >0.1 pg of each allele in the <u>total</u> reaction. $C_u$ is defined by the number of cycles required to give <1000 pg total product <u>per 10µl of reaction</u>. For large volume PCR reaction, $C_1$ will be unchanged but $C_u$ will be somewhat larger than with a 10µl reaction.

The window of amplification cycles which generates an appropriate ammount of amplification product for characterisation (0.1-4000 pg of product) is for example 10-15 cycles for 100 ng of genomic DNA, 18 cycles for 1ng and 25 cycles for single cell amplification (6pg). The number of PCR cycles may need to be increased to detect larger alleles which amplify less efficiently. Depending on allele length 1000-$10^6$pg of faithfully amplified product can be obtained.

A particular advantage of the claimed method is that differences of as little as one repeat unit in an informative region may be identified. The method of the present invention is also believed to offer faithful reproduction of large informative genetic loci, for example those of up to 15kb, such as up to about 10kb.

It has also been found that the problems associated with the networking of amplified copies of sample DNA as well as the generation of incomplete extension products can lead to the appearance of significant amounts of single stranded minisatellite DNA of heterogeneous size. This single stranded DNA can produce significant levels of aspecific products in the PCR reaction. This problem may be overcome or at least alleviated by the use of enzymes which specifically digest or degrade single stranded DNA and which leave double stranded DNA intact. The preferred enzyme is S1 nuclease. Digestion of the final PCR products with such an enzyme results in a cleaner profile of PCR products revealed at the final detection step. Therefore in a preferred aspect of the present invention the method of the invention includes the use of at least one enzyme which

specifically digests or degrades single stranded DNA whilst leaving double stranded DNA intact, whereby to ameliorate the problems of aspecific products of the PCR reaction.

Detection of the amplified products may be effected by any convenient means. The amplified products may be identified and characterised, for example using gel electrophoresis and followed if desired by hybridisation with probes hybridisable to the informative loci contained therein followed by for example autoradiography where radiolabelled probes are used. Convenient procedures are disclosed in European Patent Application, Publication No. 238329. Alternatively more direct methods may be used. Separation of the amplification products on a gel may be followed by direct visualisation thereof. Direct staining with for example ethidium bromide may be effected where a sufficient quantity of product is available. The flanking nucleotide sequences may alternatively carry a label or marker component and such label or marker may then be detected using any convenient method. Such labels or markers may include either radioactive and non radioactive components but the latter are preferred.

The number of informative regions which can be amplified simultaneously is not believed to be limited, other than by practical limitations. For example up to twenty regions could be amplified simultaneously, conveniently ten regions and in particular up to eight, seven, six, five or four regions. In a preferred aspect six regions are amplified simultaneously.

The flanking polynucleotides of the present invention may conveniently be provided in a diagnostic kit. A further aspect of the invention therefore relates to a kit comprising two complementary flanking polynucleotides capable of hybridisation to sample DNA regions adjacent an informative locus together with a control sample of DNA and instructions for their use. Examples of convenient and preferred probes include those outlined elsewhere in the specification. The kit may also include reagent nucleotides for extension of the flanking

polynucleotides across the informative genetic locus and/or extension promotor such as an enzyme.

Examples of flanking sequences to informative genetic regions which may be used in the method of the present invention are set out on the following pages. Also disclosed are details of a panel of single locus probes which are capable of hybridisation to informative genetic regions. This panel of probes was disclosed by Ray White at a meeting in Quantico, Virginia in May 1988. Polynucleotide primers capable of hybridisation to regions adjacent to such informative genetic regions may be prepared as previously outlined and used in the method of the invention. Primers may be prepared in respect of any one such region or combination of regions to provide extension products of any desired panel of informative region(s).

pMS1.1b 5'flanking sequence extending into minisatellite.

```
5' GCCATTTCCATAAACACGTATCGAGTACCTAATACTACAAAGTACCATACCAGGTACTAC
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
3' CGGTAAAGGTATTTGTGCATAGCTCATGGATTATGATGTTTCATGGTATGGTCCATGATG

   ATCCATTAAGAATGTTAATAATTAATCACGCTCATTTGCCATTGATTTTAAGTTTCCATT
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   TAGGTAATTCTTACAATTATTAATTAGTGCGAGTAAACGGTAACTAAAATTCAAAGGTAA

   TTATTAGTGTCAGGTGGTTTTACCAATGGCCTCCCTTTCCTTCATGCTTACACTAAAAAA
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   AATAATCACAGTCCACCAAAATGGTTACCGGAGGGAAAGGAAGTACGAATGTGATTTTTT

   AGAACAACAGTAATGATTGTATGTCATGCTTTTCTGTGATGAGCCTTGATGTTTTTAACA
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   TCTTGTTGTCATTACTAACATACAGTACGAAAAGACACTACTCGGAACTACAAAAATTGT

   GAATTTGATAGTTTGAGACATAAAAATTTTTGAAAAACCAACTCACGTTTCCAAACAAAA
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   CTTAAACTATCAAACTCTGTATTTTTAAAAACTTTTTGGTTGAGTGCAAAGGTTTGTTTT

   GTGAAAAGGAGAGCTGGATTCCAGCCCCGCCACCCCAGCAAATTGAGAAATCACCCCTTG
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   CACTTTTCCTCTCGACCTAAGGTCGGGGCGGTGGGGTCGTTTAACTCTTTAGTGGGGAAC

   CTGTGAGTCAGTGGGT 3'
   ||||||||||||||||
                    MINISATELLITE
   GACACTCAGTCACCCA 5'
```

pMS1.1b 3'flanking sequence, starting at the Eco RV site about 50bp from the end of minisatellite

EcoRV

5' TGGATATCTCATGCATGGGAGCCCAGATTGGGTGGTCCTGGCCCTCATGGGTTGCATATG

3' ACCTATAGAGTACGTACCCTCGGGTCTAACCCACCAGGACCGGGAGTACCCAACGTATAC

CTTCTTTGCTGCCTGTGCTGTAAGAACCATGCTAAGAACAGACCTGTGCCCCTGGGCTCT

GAAGAAACTACGGACACGACATTCTTGGTACGATTCTTGTCTGGACACGGGGACCCGAGA

CAGCAGTGCATGTGCCAGGGAGCTGTGCGTGTGGAAGCAGCAGGAAAAACAGCTGTGCCA

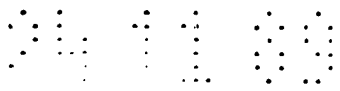
GTCGTCACGTACACGGTCCCTCGACACCCACACCTTCGTCGTCCTTTTTGTCGACACGGT

GGTCGGCGAGGCGAAGCGACAGAGAACCAGGGGCAGGGAAAGAGTTCCAAGCTCCTTGGC

CCAGCCCCTCCCCTTCCCTGTCTCTTGGTCCCCGTCCCTTTCTCAAGGTTCGAGGAACCG

CACCGNACCATGTGCTCATGGCCTCCTCCCCCGGCTGTCAGTTTCCCTATCTGTAAAATGA 3'

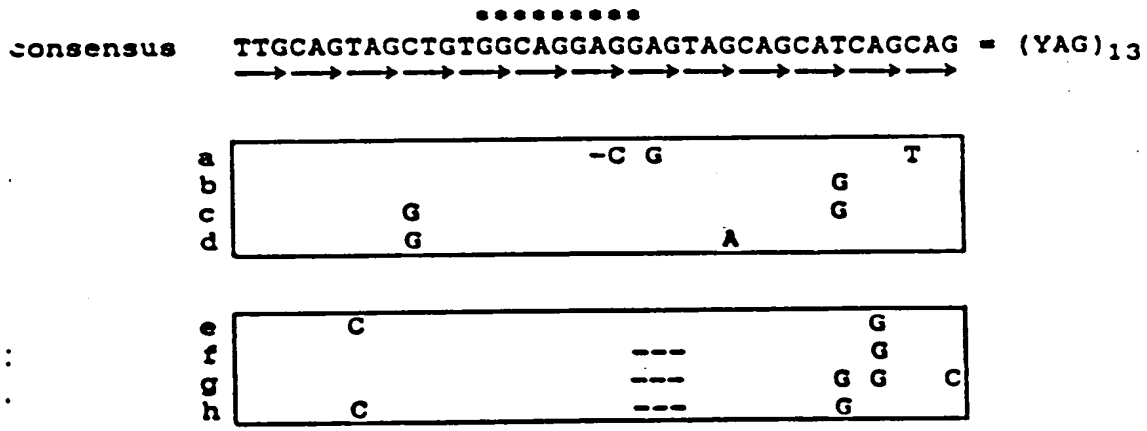
GTGGCNTGGTACACGAGTACCGGAGGAGGGGCCCGACAGTCAAAGGGATAGACATTTTACT 5'



pMS1 - the tandem repeat block is abbreviated to the two outermost repeats (in capitals) separated by X's, a novel repeat is underlined.

```
1     gatcctaaccctccaaccatgaagcccaaatctgaccccatgacccctgatttgtaaaccccccgcctggcc
73    tgtacaggtccagccccagctgctgtgccttcccccggttcagtctcctagtgaaggtgcgggagctgagca
145   atgaaggcactgttggggaaactgtagcagggaagggycaggctctatggagggtccttgatacacaccyc
217   acccccaagccttaagcagacatcaggaaaacccagtgaagggccgcaggagagatgtgtggacctgcacc
289   tgcacctccgggctctgtccatgccccaccagctacgcacttgtaccagttggtgagcgtcatcatgacgt
361   cagtgaggccagcaccaggcagaagtggaagaaggagtagctgtaggtgacgccgtcctgctcgttgtcaa
433   ggcccggcccctcacaggctgccacctgctgctgctgctgctgtggcgtctagcataggtgggcactcct
505   ggtctgcatcaggctgttcacctgccggtggtctgaggagcgcagactgcaggggagggaggaaggtgggc
577   aatagatagctctggcgttggtcataacactggacacaaatggtcccagagcctcagaggccagcgacccg
649   ttactgggcctgggagttaacaaactgctttgcctctcaactctacccctaccccccactcccatttctgt
721   tgaccttgagcaaaacacttaacctctctgagcctctgattccttgtctgtaaaataagggataatcgtac
793   tcctctagagggctgctgggggactagacggaataatgatgtaaggtctctagcatagtgcttggcacag
865   atataaattcaatcaacagtcactattgttaggattgttcctttgccatttccataaacacgtatcgagta
937   ctaatactacaaagtaccataccaggtactacatccattaagaatgttaataattaatcacgctcatttgc
1009  attgattttaagtttccattttattagtgtcaggtggttttaccaatggcctccctttccttcatgcttac
1081  ctaaaaaaagaacaacagtaatgattgtatgtcatgcttttctgtgatgagccttgatgttttaacagaa
1153  ttgatagtttgagacataaaaatttttgaaaaaccaactcacgtttccaaacaaaagtgaaaaggagagct
1225  gattccagccccgccacccccagcaaattgagaaatcacccctttgctgtgagtcagtgggtCCCTATCCAxx
1297  xxxxxxxxxxxxCTATCTACCaaataccatatggatatctcatgcatggagcccagattgggtggtcctg
1369  ccctcatgggttgcatatgcttctttgctgcctgtgctgtaagaaccatgctaagaacagacctgtgcccc
1441  gggctctcagcagtgcatgtgccagggagctgtgggtgtggaagcagcaggaaaaacagctgtgccaggtc
1513  gggaggggaagggacagagaaccaggggcagggaaagagttccaagctccttggccaccggcaccatgtgc
1585  catggcctcctccccggctgtcagtttccctatctgtaaaatgaaggtaatcaggccaggagtggtgcact
1657  ctgtaatccaaacactttggcaggctgaggtaggaggattgcttgggccaggagttcaaggccatcctggg
1729  aacatagtgtctctacaaaaaaaa
```

The minisatellite repeat sequence in clone λMS29. The region in the 39bp consensus repeat sequence similar to the 3' end of the minisatellite core sequence (GGAGGTGGGCAGGARG, Jeffreys et al,. 1985 ) is marked with asterisks. Deviations from the consensus sequence are shown for two blocks of 4 contiguous repeat units (a-d, e-h) cloned at random from λMS29. The 39bp repeat sequence consists of 13 diverged repeats of PyAG.



pMS31 5'flanking sequence extending into minisatellite

pMS31 - the tandem repeat 318 is shown in full capitals.

```
  1 gatccactcggaaccacctgcagttaggagcaagcctagaatgttctggaaggattgaagccagccttgtc
 73 gaggccctgggaaagtggcctggacatggggatgtggctggaggacccgaggaagatgctgaagtcctgtg
145 gaggcccggtctgggagccacggcccctcccccactcagtccggcctgctggggtttcctgccgggcctcc
217 tcagagcccacggctccccaggtggctctggcccgggagccacaggcacaacctaggcaggggaagccgca
289 gcacaatgttggctcttccctttgcacgctggacggtggcgttttgcctttcgccttgggctcaggagggg
361 tgggggtcaggaggggccatgaaggggacctggccttggCTGTCCACCTCCCACAGACAxxxxxxxxxxx
433 xxxTCTACCTCCCACAGACACTGCCggccggatgggcGTGTGGGGACGGTGTGCCGGTGTGGGGACGGGGT
505 CAGGTGTGGGGACGGGGTGCAGGTGTGGGGACGGGGTGCAGGTGTGGGGACGGCGTGCAGGTGTGGGGACG
577 GGTGCAGGTGTGGGGACGGCGTGCTGTGGGGATC  610
```

## FLANKING SEQUENCE 3' TO THE MS31 LOCUS

### MINISATELLITE

```
MINISATELLITE  5' CCGGCCGG ATGGGCGTGT GGGGACGGTG TGCCGGTGTG GGGACGGGGT
               3' GGCCGGCC TACCCGCACA CCCCTGCCAC ACGGCCACAC CCCTGCCCCA

     GCAGGTGTGG GGACGGGGTG CAGGTGTGGG GACGGGGTGC AGGTGTGGGG ACGGCGTGCA
     CGTCCACACC CCTGCCCCAC GTCCACACCC CTGCCCCACG TCCACACCCC TGCCGCACGT

     GGTGTGGGGA CGGGGTGCAG GTGTGGGGAC GGCGTGCTGT GGGGATC 3'
     CCACACCCCT GCCCCACGTC CACACCCCTG CCGCACGACA CCCCTAG 5'
```

5'flanking sequence of pMS32.6 extending into minisatellite

```
5' GATCACCGGTGAATTCCACAGACACTTAAAAGCAAAAATAATAATTTGTTGAATACAGTG
   ----------|----------|----------|----------|----------|----------|
3' CTAGTGGCCACTTAAGGTGTCTGTGAATTTTCGTTTTTATTATTAAACAACTTATGTCAC

   AGTTCTAAATTTCTCTTCAAAGAATCAGTATGTCAGTATGTTCAGTTCTTTGTTCTCCAT
   ----------|----------|----------|----------|----------|----------|
   TCAAGATTTAAAGAGAAGTTTCTTAGTCATACAGTCATAGGAGTCAAGAAAGCCGAGGTA

   TTTAAAGTTGAACTTCCTCGTTCTCCTCAGCCCTAGTTTCACTAAACAACCTTTTCCAAC   3'
   ----------|----------|----------|----------|----------|----------|
   AAATTTCAACTTGAAGGAGCAAGAGGAGTCGGGATCAAAGCGATTTGTTGGAAAAGGTTG   5'

   MINISATELLITE
```

pMS32.6 3'flanking sequence extending from minisatellite

MINISATELLITE

5' AACCACCCTTCCCAC
3' TTGGTGGGAAGGGTG

CAAACTACTCACCCCGCCACTCTGGCTCATACCCCTGCTCTCTTTAAAGTAGCCAATCGG
GTTTGATGAGTGGGGCGGTGAGACCGAGTATGGGGACGAGAGAAATTTCATCGGTTAGCC

AATTAGCTTAGACTGTGCAGTCCAACCCTAGCCGATACGGGAACGACGCATCAGTAGGGG
TTAATCGAATCTGACACGTCAGGTTGGGATCGGCTATGCCCTTGCTGCGTAGTCATCCCC

CTACCTGTGTCAGGAATCAGAACCCCTTCCCCTCCCTTGTTCAGGTGTGCTCTGGCCATT
GATGGACACAGTCCTTAGTCTTGGGGAAGGGGAGGGAACAAGTCCACACGAGACCGGTAA

GCTCCATCTGCGAGTCGCACCCTTCTAGAGAAGTAAAATTGCCTTGCTGAGAAAATTAAA
CGAGGTAGACGCTCAGCGTGGGAAGATCTCTTCATTTTAACGGAACGACTCTTTTAATTT

CTTATGTTTGAGTGGTATTTCTTTTGCAGCACCAAAAATTTATTTACAACAAATTCTACA 3'
GAATACAAACTCACCATAAAGAAAACGTCGTGGTTTTTAAATAAATGTTGTTTAAGATGT 5'

FMS32 - retroviral LTR underlined EI sequence in bold

```
     gatcaccggtgaattccacagacacttaaaagcaaaaataataatttgttgaatacagtgagttctaaatt
  73 ctcttcaaagaatcagtatgtcagtatgttcagttctttgttctccattttaaagttgaacttcctcgttc
 145 cctcagccctagtttcactaaacaacctttttccaacagttctaatcagtagttaacgtctattCCCCTGGT
 217 ACCTGCTCCATTCTGAGTCAxxxxxxxxxxxxxxxACCTGAATCATCCTGAGTCACCTGTTCTGTaaccac
 289 cttcccaccaaactactcaccccgccactctggctcataccccctgctctctttaaagtagccaatcggaat
 361 agcttagactgtgcagtccaaccctagccgatacgggaacgacgcatcagtaggggctacctgtgtcagga
 433 tcagaacccctttcccctcccttgttcaggtgtgctctggccattgctccatctgcgagtcgcacccttcta
 505 agaagtaaaattgccttgctgagaaaattaaacttatgtttgagtggtatttcttttgcagcaccaaaaat
 577 tatttacaacaaattctacacaaatttttccagaaactggaaatggagagcttatgttacccaattttaag
 649 tttactagaaagatacaatagacgtgacagggtggtattggtagaagaatagacacgaagatc
```

## Flanking sequences of 33.1

```
5' TGTACACTCC CAGTGCCCGA GAGGATGGGG TTGAGAGCTA AGCTGAGAAA GTCTATCTCT
3' ACATGTGAGG GTCACGGGCT CTCCTACCCC AACTCTCGAT TCGACTCTTT CAGATAGAGA

   GAATTTCANT GCTGATAAAA ACAACCGAGC CTGCCGGGGC AGGGGCTTGC TTTCTCCACG
   CTTAAAGTNA CGACTATTTT TGTTGGCTCG GACGGCCCCG TCCCCGAACG AAAGAGGTGC

   GATGGGATGC CACAACTGC
   CTACCCTACG GTGTTGACG

                        MINISATELLITE

                                 TTTCCTGAA CATCGTACCC CCCACCCCCC AGAAGCTTGT
                                 AAAGGACTT GTAGCATGGG GGGTGGGGGG TCTTCGAACA

   GGGTGACACG GCACGGTGGC TTAGGGACAC AGATGATAGA AGCTGGCTGC ATAAAAGACA
   CCCACTGTGC CGTGCCACCG AATCCCTGTG TCTACTATCT TCGACCGACG TATTTTCTGT

   AGTCTTTACA GAGTCTTGTC AAGCCCTCAG GTCTGCGCAC GTCTGATCTA ACCAAAGTTT
   TCAGAAATGT CTCAGAACAG TTCGGGAGTC CAGACGCGTG CAGACTAGAT TGGTTTCAAA

   TTCCTGTGGA 3'
   AAGGACACCT 5'
```

## Flanking sequences of 33.4

5' ATCAAGGAGG CAGTGAGAGA TGAACTGGAA GTGACCAGGG GCTGCCAGCA AACCCCCTCC
3' TAGTTCCTCC GTCACTCTCT ACTTGACCTT CACTGGTCCC CGACGGTCGT TTGGGGGAGG

ACCGAGAAGA TGACTTTCAC CTACTATACA GCAGAAAACC AAAAGCCAAG ATAAAAATCG
TGGCTCTTCT ACTGAAAGTG GATGATATGT CGTCTTTTGG TTTTCGGTTC TATTTTTAGC

CTGGGGAGGG GCAGGGATGG GGGACCGGGC CAGACCCCAG CTGCTGAGCA CGCGCCACCT
GACCCCTCCC CGTCCCTACC CCCTGGCCCG GTCTGGGGTC GACGACTCGT GCGCGGTGGA

MINISATELLITE

CCTAAGCAGG TTCTGGTGGC CTCCTGGCTG GGTGTAGGCA GAGGCTGGCT CCCGGAGGCC
GGATTCGTCC AAGACCACCG GAGGACCGAC CCACATCCGT CTCCGACCGA GGGCCTCCGG

GCAGGGGGGC TCTGAGAAGG GGGCGGCCTG AGGGGGAGCC CAGGACAGCC CCTATGCTGC
CGTCCCCCCG AGACTCTTCC CCCGCCGGAC TCCCCCTCGG GTCCTGTCGG GGATACGACG

CCCCGTCCAG CCCGGCCCCT CAGGCTGTGT TTCCTGAGAC CTCGCTAGTC TAACCCAAAC
GGGGCAGGTC GGGCCGGGGA GTCCGACACA AAGGACTCTG GAGCGATCAG ATTGGGTTTG

CTCCCAGCTC TGTCATTCAA AGGGATGGAT TAGGTCCCCG TGGGTCCATT TCCCT 3'
GAGGGTCGAG ACAGTAAGTT TCCCTACCTA ATCCAGGGGC ACCCAGGTAA AGGGA 5'

Flanking sequences of 33.6

5' TGTGAGTAGA GGAGACCTCA CATTTGACCT TGGAAAGT
3' ACACTCATCT CCTCTGGAGT GTAAACTGGA ACCTTTCA

MINISATELLITE

GGTTG CTCCTCACTC TGTGGTCTTT GTCTGTCCAG ACCTTCCCTT
CCAAC GAGGAGTGAG ACACCAGAAA CAGACAGGTC TGGAAGGGAA

CTTGGATTGA GTAATGTCTC ACCTGCTTGC 3'
GAACCTAACT CATTACAGAG TGGACGAACG 5'

pMS43 - Alu sequences are in bold, and short tandem repeats are underlined. The minisatellite arrays are abbreviated to the two outermost repeats (in capitals) separated by X's, sequencing 3' of 43A resumes at a Sma I site within the last repeat unit, and so only these 6 bases are shown in capitals.

```
1       gatctatacatgtTTACACACATGCCACACACCCTTCCCAAGGCCGTCCCTATACCCAxxxxxxxxxxxxx
73      xCCCGGGggccatcggcgcagtggcctcagcgtccagcctccccgtcctccggagctctctctggtgggggg
145     cccagtcctcctcgtcttccggagctctctctctctggtgggggccccacctggccgacctgaaggtcagg
217     catccactctttggggaaacgaagttgggttcgtaggggcagctctggtgttgttggtccttccctccaga
289     accacgtctcttcgtctggctcccaggagaccccagttctatttctatttctcccgcctctgtcatcccct
361     aaactgcttctgcatagcccccactctgcctctcaggctttaaccccaactgtgagccccacctggctgtc
433     tggggccactgaactccacgtgtccatcgtggctctccccaacatgccaccacgcaccggctggcattctg
505     ccctatgaagccctgcctgcctggggactgggaatcagccgctgtcccagcccacccgtcaatcacccatc
577     atcaccggccaatcacccgtcaatcaccgtctcttgccagctttgcatcctaaatatttttttggaagagtc
649     cgtgtccgccacaccccacaagaaagaaagttcaccctactgggctcaacagtaaagatgttaatgaaggg
721     agcttagggatgtgtggttaggggcatagggacagacgggggcagtgcaacctcaggggcggcaacgtgca
793     ttctcgacaccggcagagtgaggcctcgggcgccactcggaccacggtgcagcttccaacaacctgtggtg
865     ctccacgggccaagacagcctgcgtgcctccacgggccaagagcacaaggtgggaggagggaaggagctgg
937     ggactggaaagcggtggtcagcatctgagggcccggcacgTGGGGCTAGACGGGGGATGTGGTGAxxxxxx
1009    xxxxxxxCGGGGGCCTGGTGATGGGGacaggtgagtctcctgtccctcaccaggagaactgggttctggt
1081    ctcagcctgtctcccaccgtagagcaacctctgaccgccagacaggacaccaaggtgaaaactttcaagcc
1153    ccaggacgggagagaggtcaggggngaggtcggggnngaggtggcagggagggcagggaggtcagggagt
1225    gagtaaacgctggcgcggccacagctgtgccttgtcaggggatgaaggtgaaagacagaacccgcggtgac
1297    caggctcccctgccccgcggaggagtcaggccatcagaaaggcccctatgtcagccaggcgtggtggttca
1369    acctggaatcccagcactgtgggaggcagaggcgggcagatc
```

## FLANKING SEQUENCES OF THE MS51 LOCUS

5' GATCAGCGAA CTTCCTCTCG GCTCCCGATA TCCTCCTCGA TACGCACTCT GCCACAACGG
3' CTAGTCGCTT GAAGGAGAGC CGAGGGCTAT AGGAGGAGCT ATGCGTGAGA CGGTGTTGCC

GCAGGGTCCC TTTCAGCGTC TCATCCACAG TGAACGGGAG TTGAGGCTTT CTTAGCGGAG
CGTCCCAGGG AAAGTCGCAG AGTAGGTGTC ACTTGCCCTC AACTCCGAAA GAATCGCCTC

GGGCTGGAGG GAC
CCCGACCTCC CTG

## MINISATELLITE

CCTC CCTGCCGTTT CCGGATGCTA CGGGGTGGAT CGGAGTGTGG
GGAG GGACGCCAAA GGCCTACGAT GCCCCACCTA GCCTCACACC

TGTTAAGCAC ATCTGGACAC GCTCTGTCCG AGACACATAG TCCCCAGGCG ACCTACAGCC
ACAATTCGTG TAGACCTGTG CGAGACAGGC TCTGTGTATC AGGGGTCCGC TGGATGTCGG

ACAGCCTGAC CTCCTGAAAA TTTCCCAGCT TCCCACAGTC CTCAATGTGG AAACCAGTGC
TGTCGGACTG GAGGACTTTT AAAGGGTCGA AGGGTGTCAG GAGTTACACC TTTGGTCACG

CCAAAGGCCA CCTGCCCACA CTGGCACCGA ATTC 3'
GGTTTCCGGT GGACGGGTGT GACCGTGGCT TAAG 5'

MS51- Sequence across the adjacent Hinf I site

```
  1  AGTGGATTTT CCCACTTCCC TCTGTTGTCT ATAATGAATG ACCACTACAT
 51  TTTTAACCCC AAAAGCTTCT CATATGAAGA GGAGACTTCT TTCAGCAGCA
101  GGCAGGGCAC CTTCTCTGAA GCTCGCCAGC AAGCACCAAC CGACTCCCTG
151  GCCACGGCAG CTGACGCCCA GGCAGTGACC CCACACGGGA GGGCTCTCGA
201  AAAGRCTGGA GCTCCCATGA C
```

pMS228 - sequnce from a 2kb subclone of DNA surrounding 228B. Alu sequences are in bold, and short tandem repeats are underlined. The minisatellite arrays are abbreviated to the two outermost repeats (in capitals) separated by X's.

```
1    aggccacccccaccaacaacataaccagaggggaaggaagtcaggccccttctcactctgagaagctggtg
73   ctgggatttttaggctgtcacgacgattccacccggccagggcaggcccgaaccggccggaggccacaggag
145  accaatgagcctggctggactcctgcaaaccttgagggacgccgagattcacattcactgaattttcatgc
217  acgtgacactcttcttcttttgttccctcccgccctgccccacggagccatttacaaacataaaaccattt
289  taaaccatttttaaatggtttaaaacctgtttctgtaccaagtggtacagaaatagggccagcccccggt
361  cagcgccacgagctcctagggccagagtgcaagagaggcACAGGGCGAGAGGGGGxxxxxxxxxxxxxxx
433  GAGAGGGGGACAGGGtcatcagggtgcttagggtgggctccggggcgtctgcaccaccaggcgcacagccc
505  ggaggtggcaggagtcatccgttctgaaacagccagaggtacaacctcgtcgtccaggcaccggccgagtt
577  ggactcagggtcaaagccaagctgaggcaacgtcgagatggagggtaacagccctcagcctgcacctgcca
649  actgcggaggccccacaggaacaatccgggagggtggggtggtgcctgcctggcagctgcgggggctgggt
721  gggaagggctactccaccctggaggcccagctcacaccaacctcctagcccctgacgtcccaccaggcagc
793  tcacaaggttacaggtcggttccttctccactggattcctcccacatcgggtgacctgaccacacacacgg
865  aggtgcccagcggtggtcccagccccaacatctcaagagcaggacacccgagtggagatactaggtcacag
937  aatgtctccacacagacattcaggcaggttcgagggaagaagacagcttcccggccactctcccaccacgc
1009 acacccggtgggctcctctcctcaacctgggcccacattctctcccaggttactcacatcactcagtcatc
1081 ctcacatcactcacatcgccccatgacatccgcctctgagctgccagcctcccctccccagccccttttcttc
1153 ttccctcccctccctccctcctttctttttttttgaaacaggtcacccaggctggagttcagtggcacaatct
1225 gactcactgcagcctccgcctctgggctcaagccttccaggctcaagcaatcctcagcctcagcctccca
1297 gtagttggaaacgtaactgggcaccaccatgcccagctatttttttttttttttcagtagagatgaggtctc
1369 ctacattacccaggctggtctcagactcctggtctcaagcaatcttctcaccttggcctcccaaagtgcta
1441 gattacaggtgtgagccactgcgcccgacttccccagccccttttctgacccacagcctgggatc
```

Flanking sequences of the hypervariable region 5' to the human insulin gene
(AM. J. HUM. GENET., 1986, 39, 291-229.)

```
5' CTGGGGCTGC TGTCCTAAGG CAGGGTGGGA ACTAGGCAGC CAGCAGGGAG GGGACCCCTC
3' GACCCCGACG ACAGGATTCC GTCCCACCCT TGATCCGTCG GTCGTCCCTC CCCTGGGGAG

   CCTCACTCCC ACTCTCCCAC CCCCACCACC TTGGCCCATC CATGGCGGCA TCTTGGGCCA
   GGAGTGAGGG TGAGAGGGTG GGGGTGGTGG AACCGGGTAG GTACCGCCGT AGAACCCGGT

   TCCGGGACTG GGG          MINISATELLITE                        ACAG
   AGGCCCTGAC CCC                                               TGTC

   CAGCGCAAAG AGCCCCGCCC TGCAGCCTCC AGCTCTCCTG GTCTAATGTG GAAAGTGCCC
   GTCGCGTTTC TCGGGGCGGG ACGTCGGAGG TCGAGAGGAC CAGATTACAC CTTTCACCGG

   CAGGTGAGGG CTTTGCTCTC CTGGAGACAT TTGCCCCCAG 3'
   GTCCACTCCC GAAACGAGAG GACCTCTGTA AACGGGGGTC 5'
```

FLANKING SEQUENCES OF 5' ALPHA GLOBIN HVR
(AM. J. HUM. GENET., 1988, 43, 249-256)

```
5' TTAGATGCTC GCCG    MINISATELLITE    CTGTGCGGG AAGCCCGAAA TCCTTA 3'
3' AATCTACGAG CGGC                     GACACGCCC TTCGGGCTTT AGGAAT 5'
```

FLANKING SEQUENCES OF THE ALPHA GLOBIN 3' HVR
(EMBO, 1986, 5, 1957-1863)

5' AGTCCCACCT GCAGGAAAAG GTGCAGGTAA GG
3' TCAGGGTGGA CGTCCTTTTC CACGTCCATT CC

MINISATELLITE

AGGAACAGCA ACACGCAGGG AATATCGACA TGGGTGATGC
TCCTTGTCGT TGTGCGTCCC TTATAGCTGT ACCCACTACG

CTGCAAAGCA CAGCATCAAT CCCAAGTGAC TGA 3'
GACGTTTCGT GTCGTAGTTA GGGTTCACTG ACT 3'

Planking sequences of hypervariable region at the Zeta globin locus
(PNAS, 1983, 80, 5022-5026)

```
5' ACACCCATCA ATGGGAGCAC CAGGACAGACAT GGAGGCTAAT GTCATGTTGT AGACAGGAT
3' TGTGGGTAGT TACCCTCGTG GTCCTGTCTGTA CCTCCGATTA CAGTACAACA TCTGTCCTA

   GGTGCTGAGC TGACCACACC CACATTATTA GAAAATAACA GCACAGGCTT GGGGTGGAGG
   CCACGACTCG ACTGGTGTGG GTGTAATAAT CTTTTATTGT CGTGTCCGAA CCCCACCTCC

   CGGGACACAA GACTAGCCAG AAGGAGAAAG AAAGGTGAAA AGCTGTTGGT GCAAGGAAGC
   GCCCTGTGTT CTGATCGGTC TTCCTCTTTC TTTCCACTTT TCGACAACCA CGTTCCTTCG

   TCTTGGTATT TTCAACGGCT                MINISATELLITE              AGC
   AGAACCATAA AAGTTGCCGA                                           TCG

   TACAGGGAGA AAAGACTTGG TGCTGTGGGC CTGCCTTGGG GCTGGTGGTA CAGCCCTTAT
   ATGTCCCTCT TTTCTGAACC ACGACACCCG GACGGAACCC CGACCACCAT GTCGGGAATA

   CTGCTGCCCT CAGGATCTCC CGGCCCCTCT CGTCCAGGCC CCTGCAACCC CATGCCCCAG
   GACGACGGGA GTCCTAGAGG GCCGGGGAGA GCAGGTCCGG GGACGTTGGG GTACGGGGTC

   CCTCTGAGGA CCAAAGCGCC CCCTGCTTGG GAAGAGGGGG CTCAGGGGAG TCGCCTGACC
   GGAGACTCCT GGTTTCCGCG GGGACGAACC CTTCTCCCCC GAGTCCCCTC AGCGGACTGG

   CGGTTCCAAG CCAGGCTGAT TTACCGTTGT TAACATCCTA GTGCACGCAT CCCTCTGCCT
   GCCAAGGTTC GGTCCGACTA AATGGCAACA ATTGTAGGAT CACGTGCGTA GGGAGACGGA

   CATGCACCCA ACTCCAAGGC CTGGTACAC 3'
   GTACGTGGGT TGAGGTTCCG GACCATGTG 5'
```

FLANKING SEQUENCES OF HA RAS LOCUS
(NATURE, 302, 33-37)

```
5' TGCTCCAAGG GGCTTGCCCT GCCTTGGGCC AAGTTCTAGG TCTGGCCACA GCCACAGACA
3' ACGAGGTTCC CCGAACGGGA CGGAACCCGG TTCAAGATCC AGACCGGTGT CGGTGTCTGT

   GCTCAGTCCC CTGTGTGGTC ATCCTGGCTT CTGCTGGGGG CCCACAGCGC CCCTGGTGCC
   CGAGTCAGGG GACACACCAG TAGGACCGAA GACGACCCCC GGGTGTCGCG GGGACCACGG

   CCTCCCCTCC CAGGGCCCGG GTTGAGGCTG GGCCAGGCCT CTGGGACGGG GACTTGTGCC
   GGAGGGGAGG GTCCCGGGCC CAACTCCGAC CCGGTCCGGA GACCCTGCCC CTGAACACGG

   CTGTCAGGGT TCCCTATCCC TGAGGTTGGG GGAGAGCTAG CAGGGCATGC CGCTGGCTGG
   GACAGTCCCA AGGGATAGGG ACTCCAACCC CCTCTCGATC GTCCCGTACG GCGACCGACC

   CCAGGGCTGC AGGGAC              MINISATELLITE
   GGTCCCGACG TCCCTG

   GAGTCACCAG CTTCCCCATC GATAGACTTC CCGAGGCCAG GAGCCCTCTA GGGCTGCCGG
   CTCAGTGGTC GAAGGGGTAG CTATCTGAAG GGCTCCGGTC CTCGGGAGAT CCCGACGGCC

   GTGCCACCCT GGCTCCTTCC ACACCGTGCT GGTCACTGCC TGCTGGGGGC GTCAGATGCA
   CACGGTGGGA CCGAGGAAGG TGTGGCACGA CCAGTGACGG ACGACCCCCG CAGTCTACGT

   GGTGACCCTG TGC 3'
   CCACTGGGAC ACG 5'
```

## Panel of Single Locus Probes Proposed by Ray White

| | | | | Number of Alleles | Size Range | Range of Allele Frequencies | Probability of 2 Individuals Matching | Prob. of 2 Sibs Matching | Exclusion* | |
|---|---|---|---|---|---|---|---|---|---|---|
| D17530 | YNZ22 | Zeta globin | HinfI | 14 | 0.5-1.3kb | 0.0033-0.29 | 0.0377 | 0.33 | 70% | NAR 16,5705 |
| D2544 | YNH24 | HBV-2 | HinfI | >30 | 1-5kb | 0.033-0.05 | 0.001 | 0.26 | 95% | NAR 15,10073 |
| D957 | pEFD126.3 | HBV-4/5 | HinfI | 5 | 1-2kb | | 0.0349 | 0.33 | 71% | NAR 15,10607 |
| D14513 | MLJ14 | Myoglobin | HinfI | >20 | 4-15kb | 0.0046-0.096 | 0.0042 | 0.27 | 91% | NAR 16,381 |
| D19520 | JCZ3.1 | Zeta | HinfI | >10 | 1.5-4kb | | 0.0554 | 0.35 | 65% | NAR 16,1229 |
| D16583 | EKHA21 | | AluI | | | | 0.0302 | 0.32 | 74% | |
| D1574 | CYNA13 | | AluI | | | | 0.0065 | 0.28 | 88% | |
| | pEFD64.1 | HBV4/5 | HinfI | | | | 0.091 | 0.396 | 55% | |
| | | | | | | TOTAL | $6 \times 10^{-15}$ | $1 \times 10^{-4}$ | | |

* A priori probability that a falsely accused putative father will be excluded by probe

## Additional Single Locus Probes Cloned by Ray White

| | | | | | | Heterozygosity | |
|---|---|---|---|---|---|---|---|
| D1754 | pCMM86 | Myoglobin | HinfI | >10 | 5-1.3kb | 90% | NAR 16,5223 |
| D17526 | EFD52 | HBV-4 | MspI[x] | >8 | 5-10kb | 83% | NAR 16,786 |
| D17528 | pYNH37.3 | HBV-2 | MspI[x] | 5 | 2-4kb | 78% | NAR 16,782 |
| D75396 | pJCZ67 | Zeta globin | MspI[x] | 4 | 3-6kb | 80% | NAR 16,4191 |
| D13552 | cMHZ47 | YNZ22 | MspI[x] | >10 | 1.5-3.0kb | 80% | NAR 16,3119 |
| D14523 | cKKA39 | | MspI[x] | >10 | 2-4.5kb | 83% | NAR 16,3120 |

[x] Use of HinfI not investigated

REFERENCES

1. Bell G I, Selby J M and Rutter W J
(1982) Nature 295, 31-35.

2. Capon D J, Chen E Y, Levinson A D, Seeburg P H and Goeddel D V
(1983) Nature 302, 33-37.

3. Goodbourn S E Y, Higgs D R, Clegg J B and Weatherall D J
(1983) Proc. Nat. Acad. Sci. USA 80, 5022-5026.

4. Jeffreys A J
(1987) Biochem. Soc. Trans. 15, 309-317

5. Jeffreys A J, Wilson V and Thein S L
(1985) Nature 314, 67-73

6. Jeffreys A J, Wilson V and Thein S L
(1985) Nature 316, 76-79

7. Vassart G, Georges M, Monsieur R, Brocas H, Lequarre A S and Christophe D
(1987) Science 235, 683-684

8. Wong Z, Wilson V, Jeffreys A J and Thien S L
(1986) Nucleic Acids Res. 14, 4605-4616

9. Nakamura Y, Leppert M, O'Connell P, Wolff R, Holm T, Culver M, Martin C, Fujimoto E, Hoff M, Kumlin E and White R
(1987) Science 235, 1616-1622.

10. Wong Z, Wilson V, Patel I, Povey S and Jeffreys A J
(1987) Ann. Hum. Genet. 51, 269-288.

11. Reeder S T, Breuning M H, Davies K E, Nicholls R D, Jarman A P, Higgs D R, Pearson P L and Weatherall D J
(1985) Nature 317, 542-544.

12. Donis-Keller H, Green P, Helms C, Cartinhour S, Weiffenbach B, Stephens K, Keith T P, Bowden D W, Smith D R, Lander E S, Botstein D, Akots G, Rediker K S, Gravius T, Brown V A, Rising M B, Parker C, Bricker A, Phipps P, Muller-Kahle H, Fulton T R, Ng S, Schumm J W, Barker D F, Crooks S M, Lincoln S E, Daly M J and Abrahamson J
(1987) Cell 51, 319-337.

13. Nakamura Y, Lathrop M, O'Connell P, Leppert M, Barker D, Wright E, Skolnick M, Kondoleon S, Litt M, Lalouel J -M and White R
(1988) Genomics 2, 302-309.

14. Jeffreys A J, Brookfield J F Y and Semeonoff R
(1985) Nature 317, 818-819.

15. Baird M, Balazs I, Giusti A, Miyazaki L, Nicholas L, Wexler K, Kanter E, Glassberg J, Allen F, Rubinstein P and Sussman L
(1986) Am.J.Hum.Genet 39, 489-501.

16. Thein S L, Jeffreys A J and Blacklock H A
(1986) Lancet ii, 37.

17. Knowlton R G, Brown V A, Braman J C, Barker C, Schumm J W, Murray C, Takvorian T, Ritz J and Donnis-Keller H
(1986) Blood 68, 378-385.

18. Gill P, Jeffreys A J and Werrett D J
(1985) Nature 318, 577-579.

19. Gill P, Lygo J E, Fowler S J and Werrett D J
(1987) Electrophoresis 8, 38-44.

20. Mullis K, Faloona F, Scharf S, Siki R, Horn G and Erlich H
(1986) Cold Spring Harbor Symp. Quant.Biol. 51, 263-273.

21. Saiki R K, Gelfand D H, Stoffel S, Scharf S J, Higuchi R, Horn G T, Mullis K B and Erlich H A
(1988) Science 239, 487-491.

22. Higuchi R, Von Beroldingen C H, Sensabaugh G F and Erlich H A
(1988) Nature 332, 543-546.

23. Li H, Gyllensten U B, Cui X, Saiki R K, Erlich H and Arnheim N
(1988) Nature, 335, 414-417.

24. Wong C, Dowling C E, Saiki R K, Higuchi R G, Erlich H A and Kazazian H H
(1987) Nature 330, 384-386.

25. Saiki R D, Scharf S J, Faloona F, Mullis K B, Horn G T, Erlich H A and Arnheim N
(1985) Science 230, 1350-1354.

26. Jeffreys A J and Morton D B
(1987) Animal Genetics 18, 1-15.

27. Vieira J and Messing J
(1982) Gene 19, 259-268.

28. Feinberg A P and Vogelstein B
(1984) Anal.Biochem. 137, 266-267.

29. Maniatis T, Fritsch E F and Sambrook J
(1982) Molecular cloning: a laboratory manual
(Cold Spring Harbor Laboratory, N.Y.).

30. Armour J A L, Patel I, Thein S L, Fey M F and Jeffreys A J
(1988) Genomics, 4, 328-334.

31. Jeffreys A J, Royle N J, Wilson V and Wong Z
(1988) Nature 332, 278-281.

32. Royle N J, Clarkson R, Wong Z and Jeffreys A J
(1988) Genomics, 1, 352-334

33. Royle N J, Clarkson R, Wong Z and Jeffreys A J
(1987) Human Gene mapping 9: Ninth international workshop on human gene mapping, Cytogenet, Cell Genet. 46, 685.

The invention will now be illustrated but not limited by reference to the following:

Example 1

MATERIALS AND METHODS

(a) Preparation of genomic DNA, oligonucleotides and hybridization probes

Human DNA samples were provided by CEPH, Paris, or were prepared from venous blood as described elsewhere [26]. Oligonucleotides synthesised on an ABI 380B DNA synthesiser using reagents supplied by Cruachem were purified by ethanol precipitation and dissolved in water. The 5.6 kb Sau3A insert from human minisatellite clone MS32 [10] was subcloned into the BamHI site of pUC13 [27]. Similarly, the minisatellite inserts from recombinant M13 RF DNAs 33.1, 33.4 and 33.6

[5] were isolated as a 1.9 kb BamHI-EcoRI fragment, a 2.7 kb Sau3A-EcoRI fragment and a 0.7 kb BamHI-EcoRI respectively, and subcloned into pUC13 digested with BamHI plus EcoRI, to produce the plasmid series p33.1, p33.4 and p33.6. Appropriate minisatellite-containing DNA fragments were isolated from restricted plasmid DNAs by electrophoresis through 1% low gelling temperature agarose (Sea Plaque); gel slices containing DNA fragments were dissolved in water at 65° to a final concentration of 2μg/ml DNA. 10ng aliquots of DNA were labelled with $^{32}P$ by random oligonucleotide priming [28].

(b) Polymerase chain reaction

Aliquots of human DNA, diluted if necessary with 5mM Tris-HCl (pH7.5) in the presence of 0.1 μM oligonucleotide primers as carrier, were amplified in 10μl 67mM Tris-HCl (pH8.8), 16mM $(NH_4)_2SO_4$, 6.7mM $MgCl_2$, 10mM 2-mercaptoethanol, 6.7 μM EDTA, 1.5mM dATP, 1.5mM dCTP, 1.5mM dTTP, 1.5mM dGTP (Pharmacia), 170μg/ml bovine serum albumin (DNase free, Pharmacia) plus 1μM of each oligonucleotide primer and 1.5 units Taq polymerase (Anglian Biolabs). Reaction mixes in 1.5 ml Eppendorf tubes were overlaid with 40 μl paraffin oil and cycled for 1 minute at 95°, 1 minute at 60° and 15 minutes at 70° on an Intelligent Heating Block (Cambio, Cambridge). Final amplification reactions were generaly chased by a final step of 1 minute at 60°, to anneal any remaining single-stranded DNA with primer, followed by an extension phase at 70° for 15 minutes.

(c) Southern blot analysis of PCR products

Paraffin oil was removed from PCR reactions by extraction with diethyl ether. Agarose gel electrophoresis of PCR products, Southern blotting onto Hybond N (Amersham) and hybridization with $^{32P}$-labelled minisatellite probes were carried out as described previously [10], except that competitor human DNA was omitted from all hybridizations. Restriction digests and S1 nuclease digestion of PCR products were performed by diluting 5μl PCR reaction with 25μl restriction

endonuclease or S1 nuclease buffer [29] and digesting for 30 minutes at 37° with 3 units restriction endonuclease of S1 nuclease (BRL) prior to gel electrophoresis.

(d) Isolation and PCR analysis of single human cells

Lymphocytes were isolated by diluting venous blood with a equal volume of 1xSSC (saline sodium citrate, 0.15M NaCl, 15mM trisodium citrate, pH 7.0), layering over Histopaque-1119 (Sigma) and centrifuging at 2000g for 10 min. Cells at the interface were diluted with 3 vol 1xSSC and banded again over Histopaque. Cells were pelleted by centrifuging at 2000g for 10 minutes, washed three times with 1xSSC, with centrifugation, and resuspended in 1xSSC to $10^4$ cells/ml.

Buccal cells were isolated by diluting 0.5ml saliva with 5ml 1xSSC and centrifuging at 2000g for 10 minutes. The cell pellet was rinsed three times with 1xSSC and resuspended to $10^4$ cells/ml.

Approximately 0.1$\mu$l aliquots of the cell suspensions were pipetted onto a siliconised microscope slide and rapidly examined at 100x magnification on an inverted microscope. Droplets containing a single nucleated cell were immediately diluted with 0.4$\mu$l 1xSSC and transferred to an Eppendorf tube using a disposable tip pipette. The microscope slide was re-examined to check that the cell had been removed with the droplet.

Cells were lysed prior to PCR either by heating or by treatment with sodium dodecyl sulphate (SDS) and proteinase K [23]. In the former case, the cell droplet was diluted with 4.5$\mu$l 5mM Tris-CHl (pH7.5) containing 0.1$\mu$M oligonucleotide primers, overlaid with parafffin oil and heated at 95° for 3 minutes prior to the addition of 5$\mu$l 2x concentrated PCR buffer/primers/Taq polymerase and amplification. In the latter case, the cell droplet was mixed with 0.5$\mu$l 5mM Tris-HCl (pH7.5), 0.1$\mu$M primers plus 1_l 5mM Tris-HCl (pH7.5), 40mM dithiothreithol, 3.4 $\mu$M SDS, 50 $\mu$g/ml proteinase K [23], overlaid with paraffin oil and digested at 37° for 45 minutes. 3$\mu$l water were added

to the digest, and heated at 95° for 3 minutes to inactivate proteinase K prior to addition of 5 µl 2x PCR reaction mix as above.

(a) Selection of human minisatellites for amplification by PCR

The strategy for amplifying minisatellites is shown in Figure 1. Oligonucleotide primers corresponding to unique sequence DNA flanking the minisatellite are used to drive amplification of the entire minisatellite by Taq polymerase. Amplified alleles are detected by Southern blot hybridization with a minisatellite probe located internal to the priming sites. Six cloned minisatellites were chosen for study (Table 1). Two of them, pλg3 and MS32 [8, 10], detect highly variable loci with heterozygosities of 97% and more than 40 alleles varying in the number of repeat units. The other four minisatellites, 33.1, 33.4 and 33.6 [5] and pMS51, isolated as a Sau3A-EcoRI DNA fragment cloned from a DNA fingerprint (A J Jeffreys, unpublished data), detect much less variable loci with heterozygosities of 66-77%; the alleles are however shorter than those of pλg3 and MS32 (Table 1) and should be more amenable to amplification by PCR. The flanking sequences of pλg3, 33.1, 33.4 and 33.6 have been described previously [5, 8]; the flanking DNA of λMS32 and pMS51 was sequenced as described before [8]. All flanking DNA sequences were screened against the EMBL DNA sequence database to identify repeat elements such as Alu, and PCR oligonucleotide primers A and B (Figure 1) were designed to avoid such elements. Details of all primers and hybridization probes are given in Figure 1 legend.

(b) Fidelity and efficiency of PCR amplification of human minisatellite alleles

To determine the ability of Taq polymerase to amplify in particular long minisatellite alleles, a mixture of 0.1µg genomic DNA from each of 4 individuals, giving a total of 8 different MS32 alleles ranging in length from 1.1 to 17.9 kb, was amplified for 10-20 cycles using λMS32 flanking primers A and B, followed by Southern blot hybridization

with a minisatellite probe (Figure 2A). Using 6 minutes extension times for Taq polymerase, only the four shortest alleles (1.1-2.9 kb) were efficiently amplified. Increasing the extension time to 15 minutes, to improve the chance that the Taq polymerase would progress completely across the minisatellite, gave a marked increase in yield of the next two larger alleles (4.5, 6.6 kb), though no further improvement was seen with 30 minutes extensions. The relative yield of large alleles could also be improved by increasing the concentration of Taq polymerase (Figure 2B), allowing the detection of a 10.2 kb allele, albeit faintly. Addition of extra Taq polymerase at the 13th cycle gave only a marginal improvement in yield, and there is no evidence for a significant drop in polymerase activity during these prolonged extension times. Further experiments varying annealing temperature, extension temperature and buffer concentration failed to improve the yield of large alleles (data not shown), and all further experiments used 15 minutes extension times and high concentration of Taq polymerase (1.5 units per 10$\mu$l PCR reaction).

At low cycle numbers (10 cycles), the alleles amplified appear to be completely faithful copies of the starting λMS32 alleles, as judged by their electrophoretic mobilities (Figure 2A). At higher cycle numbers (14, 17 cycles), there is an increase in background labellings; since most of this can be eliminated by digestion with S1 nuclease (data not shown), much of this background probably arises from low levels of single-stranded templates from the previous cycle which have failed to prime, and from incomplete extension products from the previous cycles which by definition cannot prime. At high cycle numbers (20), the hybridization pattern degenerates to a heterodisperse smear, as expected since the yield of PCR product becomes so high (>400ng/ml) that out-of-register annealing of single-stranded tandem-repeated minisatellite DNA will occur during the extension phase. This will lead to premature termination of extension at a reannealed site, to spurious "alleles" arising from the extension of incomplete templates annealed out-of-register to the complementary strand of a minisatellite, and to the formation of multimolecular networks of reannealed ministaellite DNA strands.

The yields of each λMS32 allele amplified by PCR were quantified by scanning densitometry (Figure 3). PCR products from 0.1 μg genomic DNA accumulate exponentially at least up to cycle 17.

The gain in product per cycle decreases monotonously with allele length, with lower gains for 6 minutes compared with 15 minutes extension times. The gain versus allele length curves extrapolate back to a gain per cycle of approximately 2.0 for very short alleles, indicating that the efficiency of denaturation and priming at each cycle is close to 100%. Final yields of an allele can be calculated from these curves; for an allele A with gain $g_A$ per cycle present initially at n molecules, the yield after c cycles is approximately $n\,g_A^c$ molecules. The molar imbalance between alleles A and B of different lengths, arising through more efficient amplification of shorter alleles, is given by $(g_A/g_B)^c$. For example, after 10 cycles of amplification with 15 minutes extension times, the molar yield of a 1 kb allele will be 18 times higher than that of a 6 kb allele; after 25 cycles, the imbalance will be 1300-fold. This imbalance is ameliorated to some extent by the more efficient detection of longer alleles by the minisatellite hybridization probe. Nevertheless, long alleles amplified by PCR will become increasingly difficult to detect with low amounts of starting human genomic DNA and high numbers of PCR cycles.

Minisatellites pλg3, pMS51, 33.1, 33.4 and 33.6 were also tested for their ability to be amplified by PCR (data not shown). In all cases, faithful amplification of all alleles tested was observed, except for the longest (> 8kb) alleles of pλg3 which as expected failed to amplify. Again, yields of PCR product fell with increasing allele length.

(c) Fidelity of amplification of single minisatellite molecules

To test whether faithful amplification of single molecules is possible, 60 and 6 pg aliquots of human DNA, equivalent to 10 and 1 cells, were co-amplified for 25 cycles using primers for both λMS32

2. A method as claimed in claim 1 wherein the test sample of genomic DNA is subjected to restriction prior to amplification and only one primer is used in respect of each informative locus to be amplified.

3 A method as claimed in claim 1 which comprises

(i) hybridising the test sample with two primers in respect of each informative locus to be amplified, each primer being hybridisable to single strands of the test sample at a region which flanks the minisatellite sequence of the informative locus to be amplified under conditions such that an extension product of each primer is synthesised which is complementary to and spans the said minisatellite sequence of each strand of the test sample whereby the extension product synthesised from one primer, when it is separated from its complement, can serve as a template for synthesis of the extension product of the other primer;

(ii) separating the extension product so formed from the template on which it was synthesised to yield single stranded molecules;

(iii) if required hybridising the primers of step (i) with the single stranded molecules obtained according to step (ii) under conditions such that a primer extension product is synthesised from the template of each of the single stranded molecules obtained according to step (ii); and

(iv) detecting the amplification products and comparing them with one or more controls;

the method being effected such that sufficient of the desired extension product is generated to be detectable but such that the yield of extension product is inadequate to permit substantial out-of-register hybridisation between complementary minisatellite template strands.

4. A method as claimed in any one of the previous claims wherein alleles of an informative locus in the test sample are of up to 15 kilobases.

5 A method as claimed in any one of the previous claims which includes the use of at least one enzyme which specifically digests or degrades single stranded DNA whilst leaving double stranded DNA intact, whereby to alleviate the formation of aspecific amplification products

6 A method as claimed in any one of the previous claims wherein amplification is effected in buffer of reduced ionic strength and at an elevated annealing temperature whereby to alleviate mispriming.

7. A method as claimed in any one of the previous claims wherein more than one informative locus is amplified simultaneously.

8. A method as claimed in any one of the previous claims for the characterisation of one molecule of an informative locus in a genomic DNA test sample.

9. A polynucleotide hybridisable to a single strand of a test sample of genomic DNA at a region which flanks the minisatellite sequence of an informative locus so as to permit formation of an extension product which is complementary to and spans the said minisatellite sequence of the strand of the test sample and wherein the informative locus is selected from any one of the following: MS1, MS29, MS31A and MS31B, MS32, MS43A and MS43B, MS51, MS228 and MS228B.

10. A polynucleotide extension product prepared from a polynucleotide primer hybridisable to a single strand of a genomic DNA test sample at a region which flanks the minisatellite sequence of an informative locus and wherein the polynucleotide extension product is complementary to and spans the said minisatellite sequence of the strand of the test sample

11. A mixture which comprises multiple faithful copies of a polynucleotide extension product as claimed in claim 10 above

12 A mixture as claimed in claim 11 above which comprises the products of at least 3 cycles of the amplification method as claimed in one of claims 1 - 7

13. A kit comprising polynucleotide primer for amplifying the minisatellite sequence at at least one informative locus in a test sample of genomic DNA the primer being hybridisable to a single strand of the test sample at a region which flanks the minisatellite sequence of the informative locus to be amplified under conditions such that an extension product of the primer is synthesised which is complementary to and spans the said minisatellite sequence of the strand of the test sample; and the kit further comprising a control sample of DNA and instructions for use.

14. A kit as claimed in claim 13 which comprises two complementary flanking polynucleotide primers in respect of each informative locus.

DATED : 24 November, 1989.

PHILLIPS ORMONDE & FITZPATRICK
ATTORNEYS FOR:
IMPERIAL CHEMICAL INDUSTRIES PLC

David B Fitzpatrick

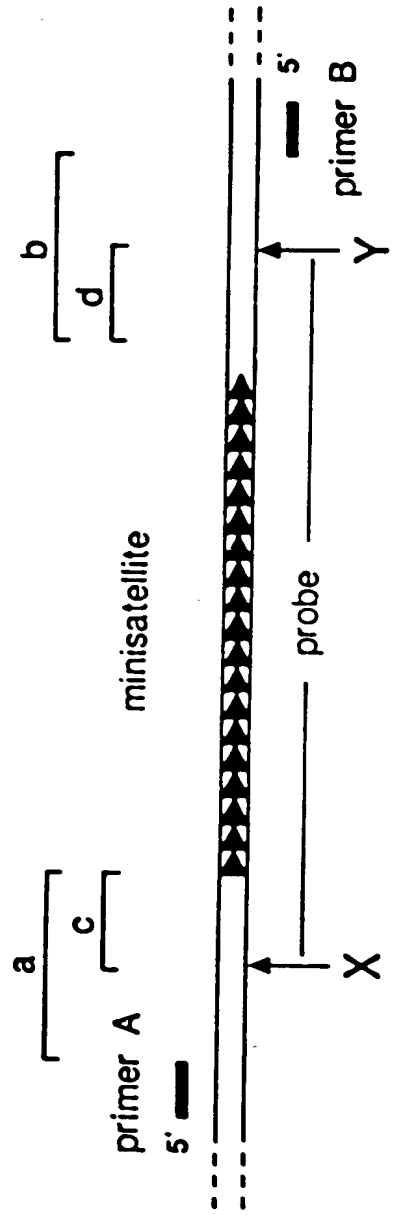



Fig. 1.

Fig.2.

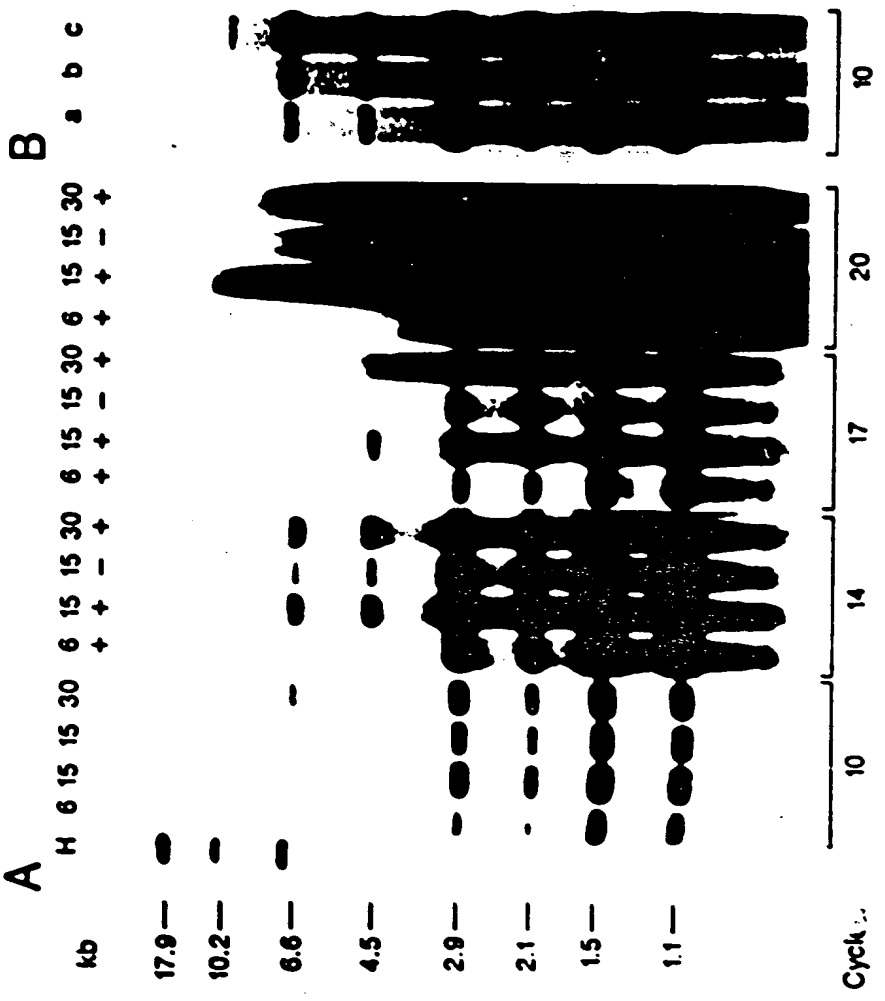
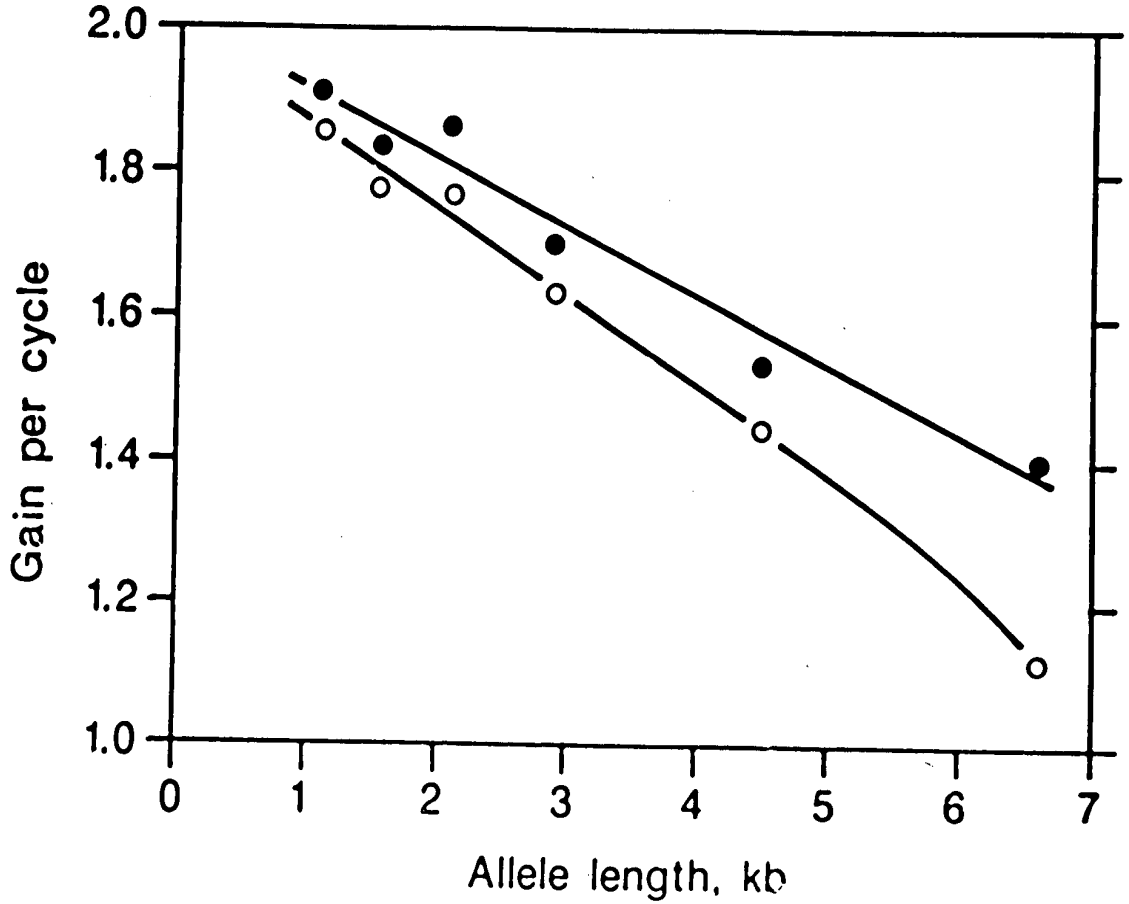


Fig. 3.



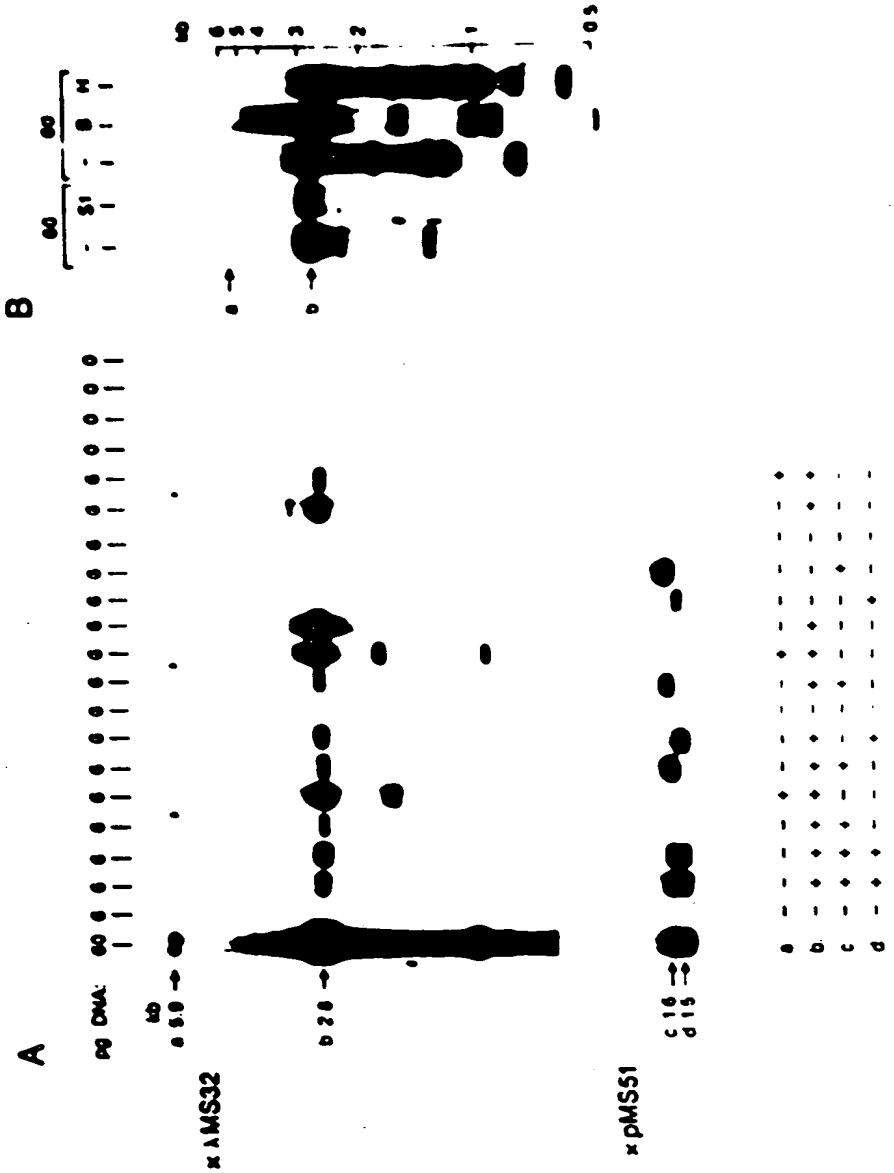



Fig. 4.


Fig. 5.

Fig. 6.

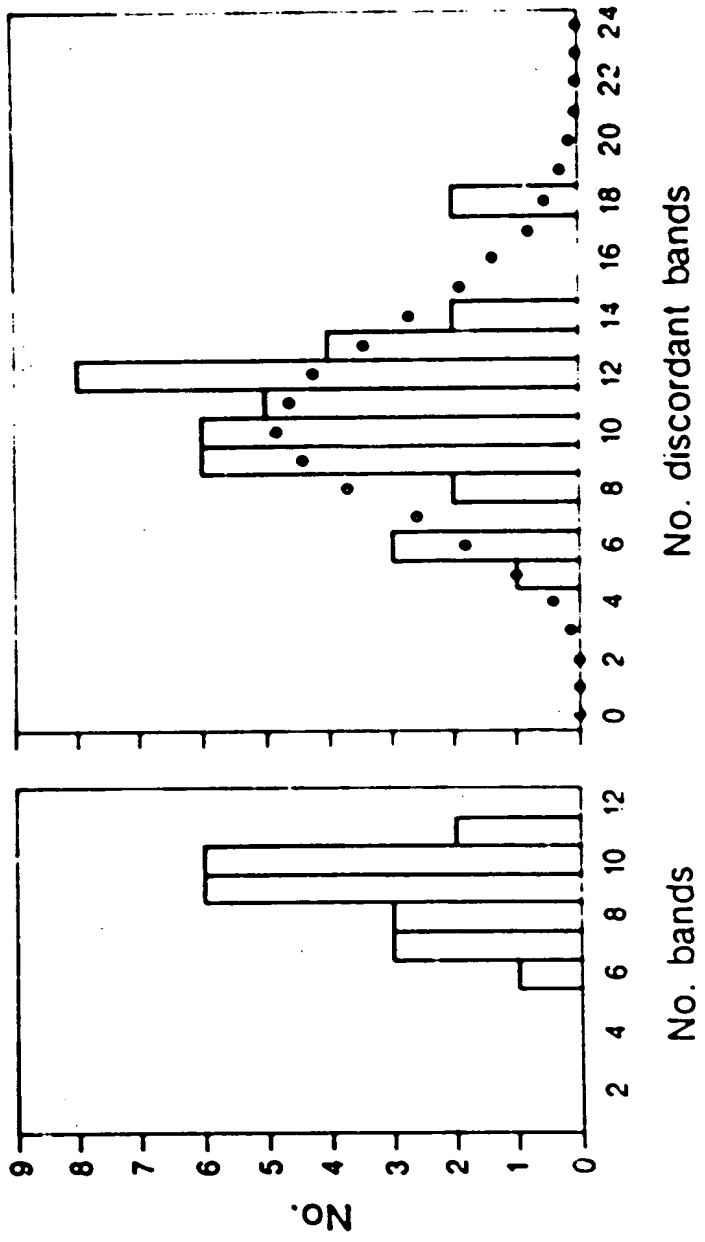


Fig. 7.

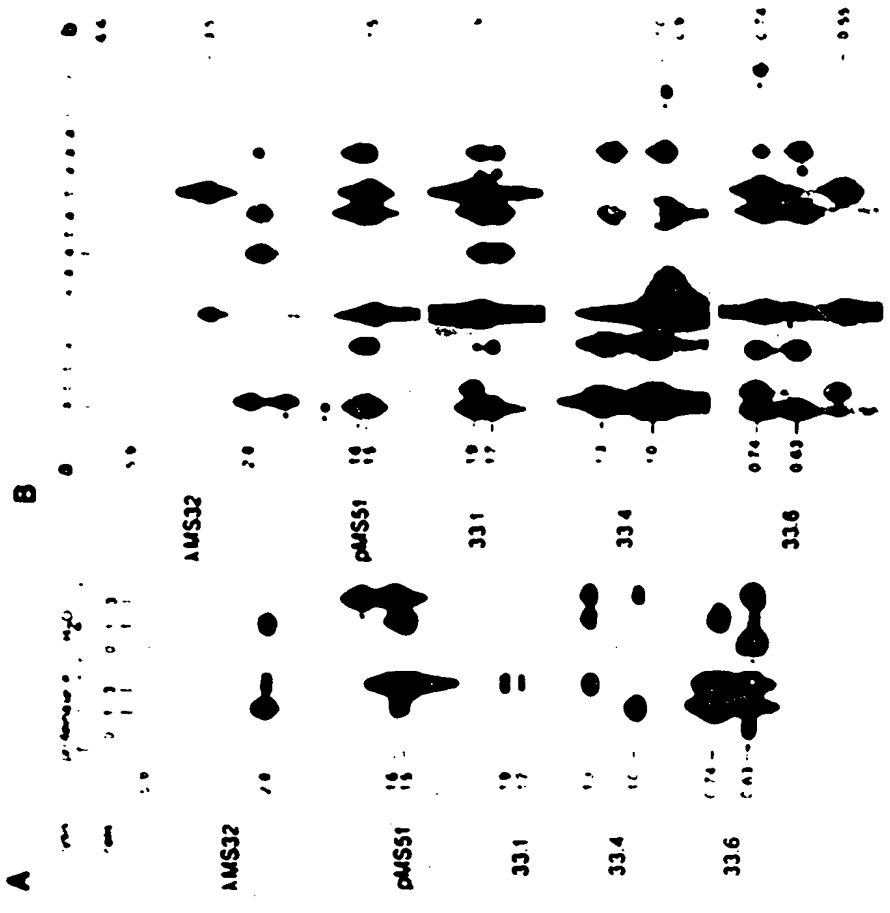


Fig. 8.

Fig.9.

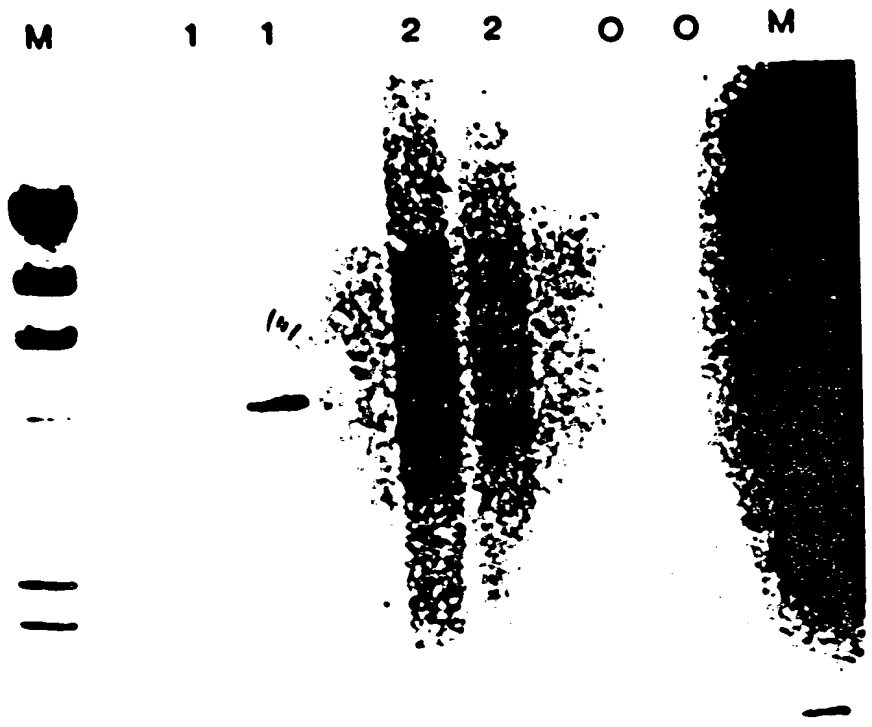
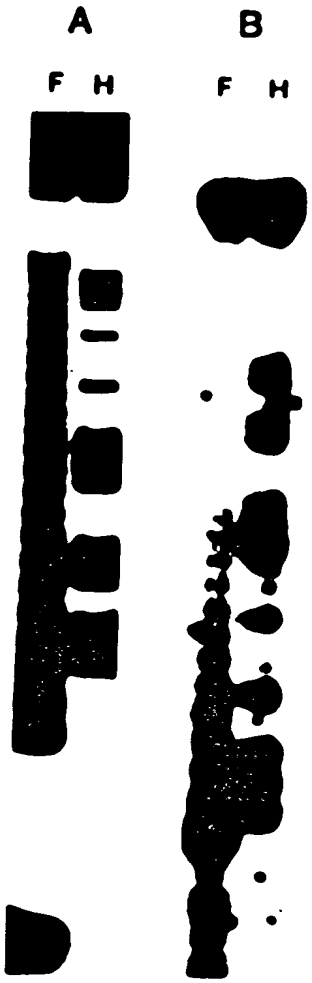


Fig. 10.



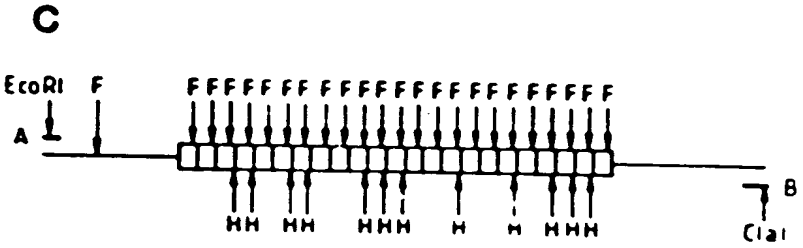
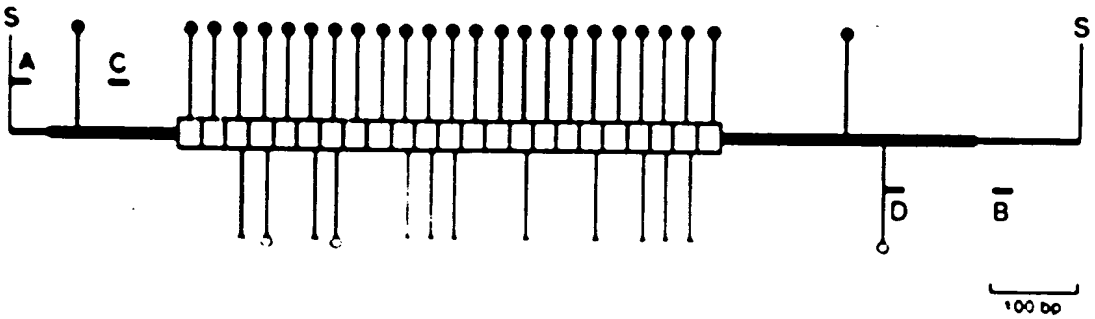
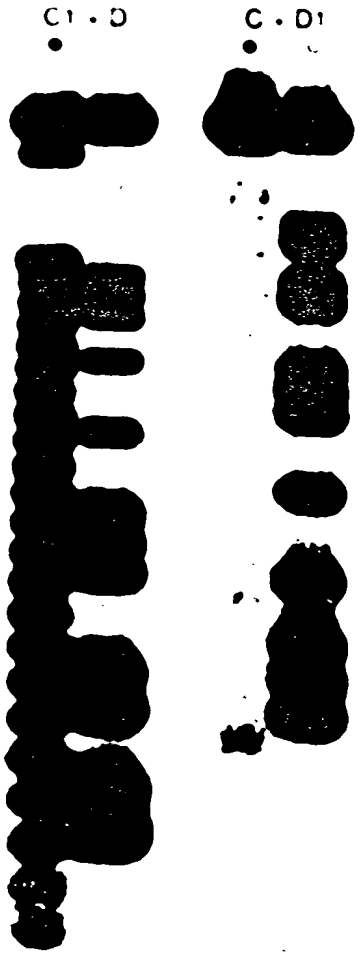



Fig. 11.

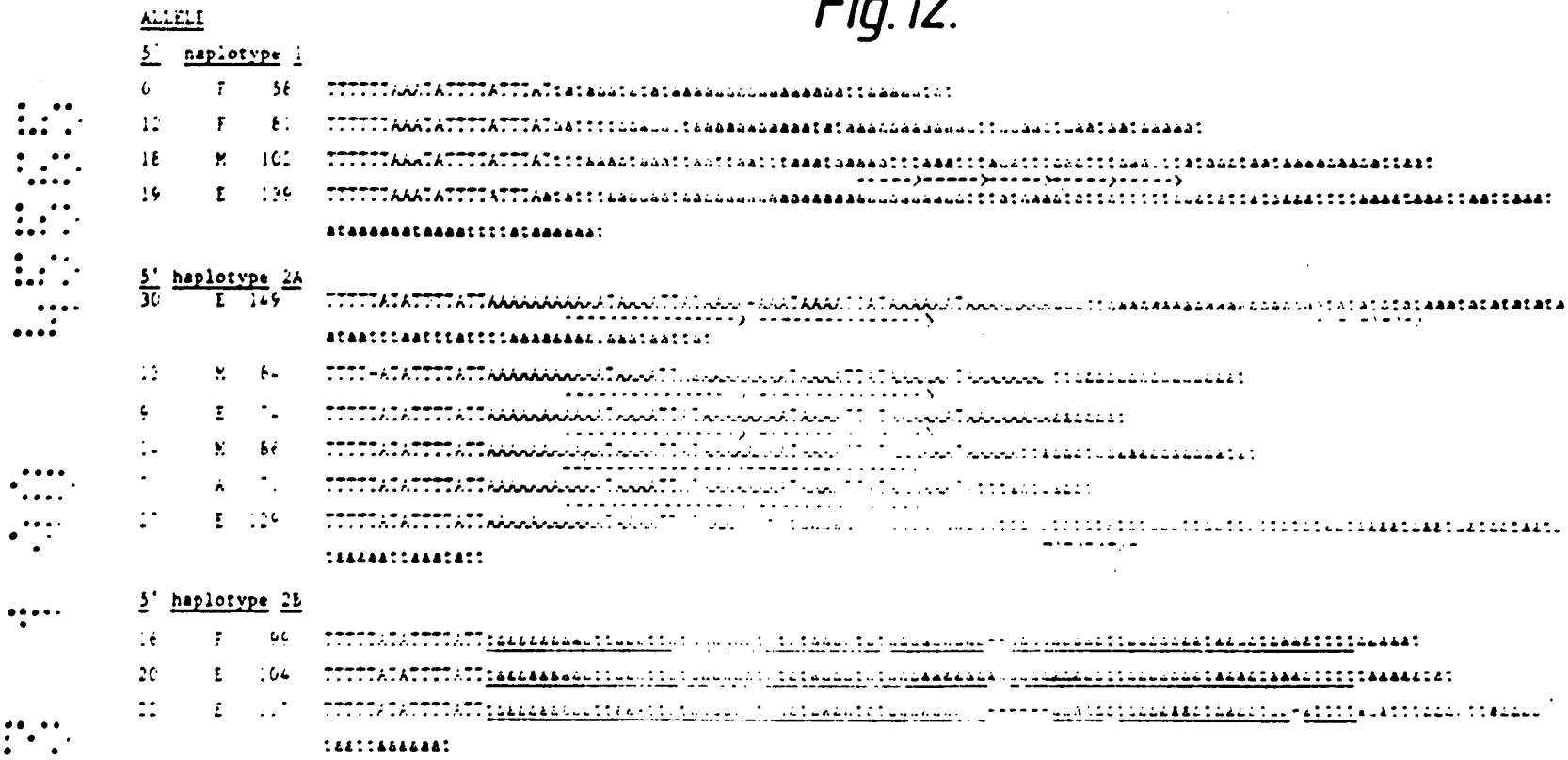
A



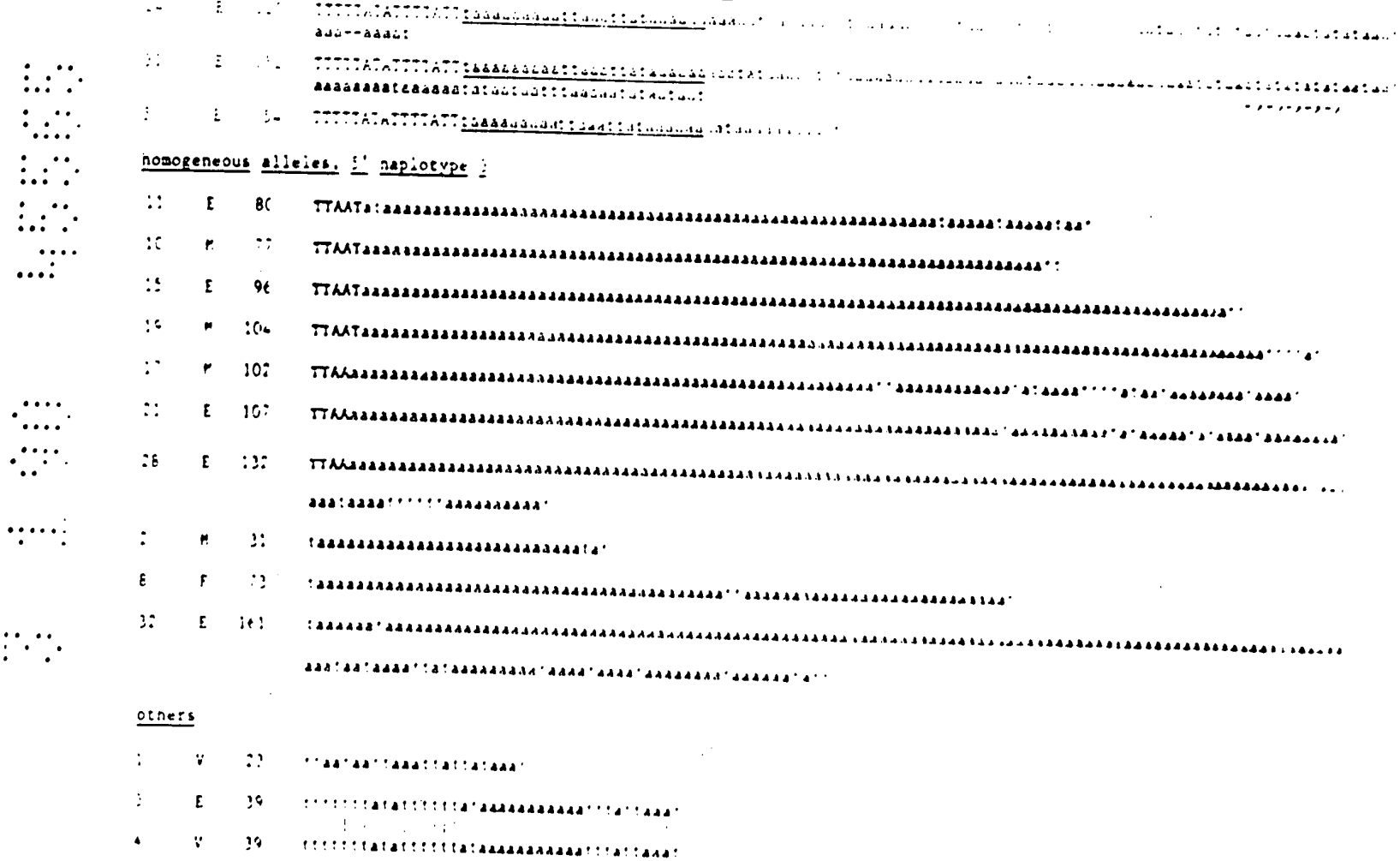

B



# Fig. 12.



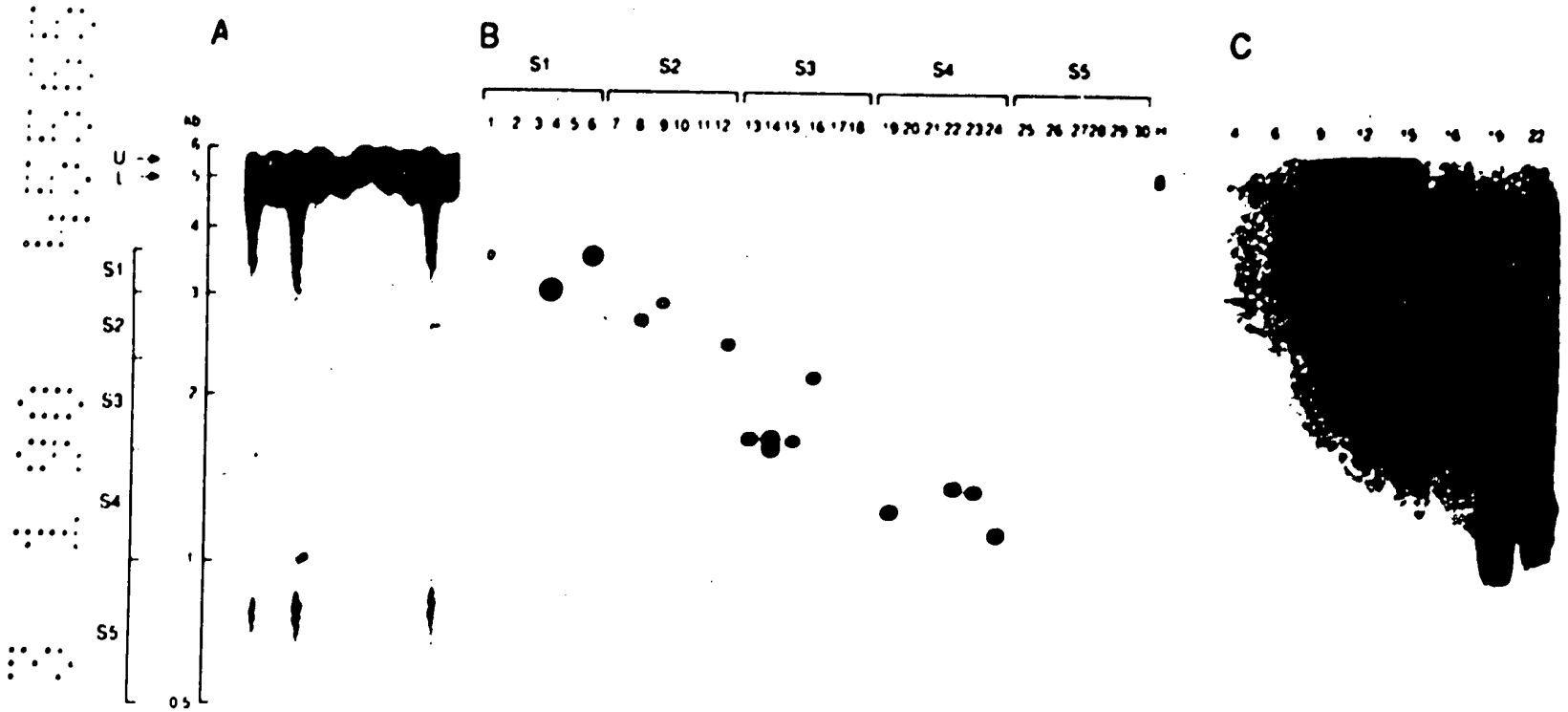
# Fig. 12 (cont.)



homogeneous alleles, 5' haplotype }

others

Fig. 13.


A
B
C
S1
S2
S3
S4
S5
U
L
kb
5
4
3
2
1
0.5
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 M